REPORT RESUMES

ED 010 363 48
LANGUAGES OF THE WORLD--INDO-EUROPEAN FASCICLE ONE.
BY- VOEGELIN, C. F. VOEGELIN, FLORENCE M.
INDIANA UNIV., BLOOMINGTON
REPORT NUMBER NDEA-VI-63-16 PUB DATE NOV 65
CONTRACT OEC-SAE-9488
EDRS PRICE MF-$0.45 HC-$12.04 301P. ANTHROPOLOGICAL
LINGUISTICS, 7(8)/1-294, NOV. 1965

DESCRIPTORS- *INDO EUROPEAN LANGUAGES, CLASSICAL LANGUAGES, ROMANCE LANGUAGES, RUSSIAN, GERMAN, FRENCH, SPANISH, PORTUGUESE, *LANGUAGES, *MODERN LANGUAGES, ARCHIVES OF LANGUAGES OF THE WORLD, BLOOMINGTON, INDIANA

THIS PUBLICATION PROVIDES A HISTORICAL REVIEW AND LISTING OF THE INDO-EUROPEAN LANGUAGES, BOTH ANCIENT AND MODERN. THE VOLUME DEALS WITH THE LESSER-KNOWN BRANCHES OF THE INDO-EUROPEAN LANGUAGE FAMILY AND THE MAIN LANGUAGES OF THIS GROUP COMMON TO MODERN TIMES. (THE REPORT IS PART OF A SERIES, ED 010 350 TO ED 010 367.) (JK)

ED010363

63-16
SAE - 9488

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Anthropological Linguistics

Volume 7 Number 8

November 1965

Title

LANGUAGES OF THE WORLD:
INDO-EUROPEAN FASCICLE ONE

A Publication of the
ARCHIVES OF LANGUAGES OF THE WORLD
Anthropology Department
Indiana University

ANTHROPOLOGICAL LINGUISTICS is designed primarily, but not exclusively, for the immediate publication of data-oriented papers for which attestation is available in the form of tape recordings on deposit in the Archives of Languages of the World. This does not imply that contributors will be restricted to scholars working in the Archives at Indiana University; in fact, one motivation for the publication of ANTHROPOLOGICAL LINGUISTICS is to increase the usefulness of the Archives to scholars elsewhere by making publishable data and conclusions and their tape recorded attestation more widely available. (Recorded attestation of papers from scholars elsewhere will be copied by the Archives and the original recordings returned to the collector; others may then work with the tape copy either in the Archives or elsewhere by having a copy sent to them.) In addition to heavily exemplified papers in the form of preliminary or final statements on restricted problems in phonemics, morphophonemics, morphemics, syntax and comparative grammar, ANTHROPOLOGICAL LINGUISTICS will include lexical lists and analyzed texts (especially in the otherwise hard-to-publish range of 20 to 100 pages) and theoretical and methodological papers (especially in the form of papers from symposia).

Each volume of ANTHROPOLOGICAL LINGUISTICS consists of nine numbers to be issued during the months of January, February, March, April, May, June, October, November and December. Subscriptions ($3.50 a year) and papers for publication should be sent to the editor, Dr. Florence M. Voegelin, Anthropology Department, Indiana University, Bloomington, Indiana.

Entered as second class matter at the postoffice
at Bloomington, Indiana

LANGUAGES OF THE WORLD:
INDO-EUROPEAN FASCICLE ONE

C.F. and F.M. Voegelin
Indiana University

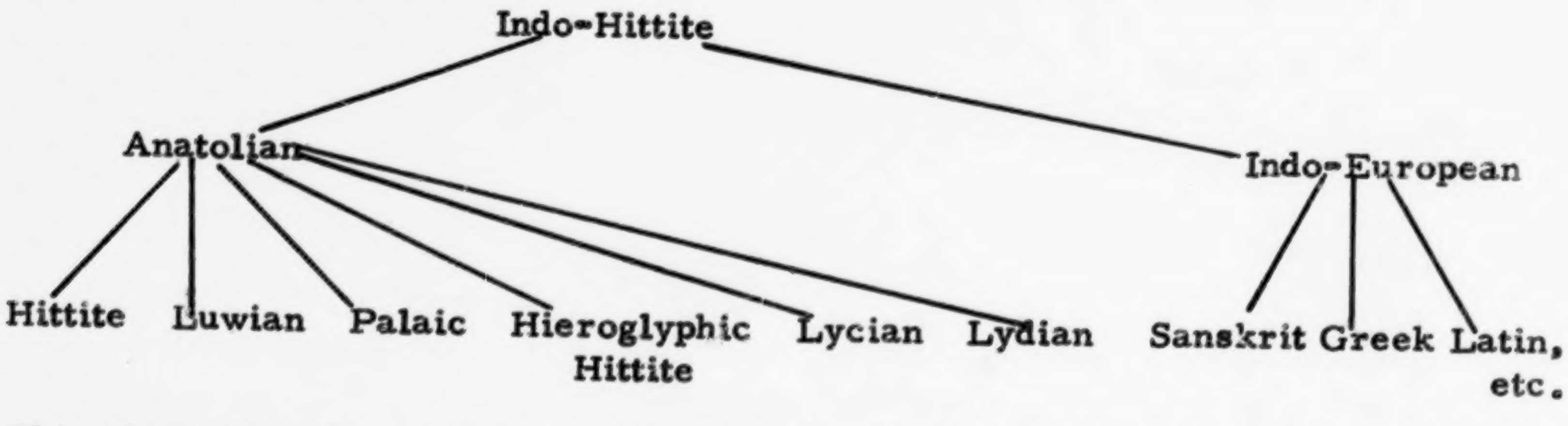
1. Branches of the Indo-European family
2. Italic
3. Germanic
4. Celtic
5. Slavic
6. Baltic
7. Greek
8. Albanian
9. Armenian
10. Iranian
11. Indic
12. Dardic

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

BRANCHES OF THE INDO-EUROPEAN FAMILY

1.0. Indo-European itself is a branch of an early language family from which it split, as is indicated below (1.1.). Subsequently, several early splits have been postulated; some of these are controversial (in conflict with each other), some not (1.2). It is possible to chart the branches of Indo-European family from the point of view of extinct and still spoken daughter languages in each (1.3). It seems simpler, however, to state some subrelationships among modern Indo-European languages in terms of their sub-branch, and others in terms of their major-branch affiliations (1.4 and 2 to 12 ff).

1.1. The following chart (after Edgar H. Sturtevant, 1947) represents the Indo-Hittite hypothesis:



This chart lists three of the so-called older Indo-European languages--the ones best preserved in writing--on one line (under Indo-European). But placing these daughter languages on one line is a matter of convenience rather than an argument in favor of the view that all three split from the parent language (Indo-European) at the same time. This parent language, however, did break away from a prior parent language (Indo-Hittite). The intermediate daughter languages--or rather the two branches after this earliest attested split--are indicated as the middle level in the chart just given (Indo-European and Anatolian).

Though Hittite offers the fullest attestation for the Anatolian branch, Hittite was not the only language in the Anatolian Branch. All have become extinct. We list them, nevertheless, to gain critical perspective on the old problem of recontructing the Indo-European homeland (culture) without reference to languages spoken in a parallel branch of an earlier language family (e.g. Herman Hirt, Die Indogermanen, ihre Verbreitung, ihre Urheimat, und ihre Kultur; 2 volumes, Strassburg, 1905, 1907).

The six ancient Anatolian languages shown in the chart above may be regarded as constituting one branch of Indo-Hittite, as Sturtevant urges; or as representing a branch of Indo-European, as other scholars have urged; or their subrelationshipp among themselves or even to Indo-European, in the usual sense, may be left unspecified beyond the cautious point of admitting that evidence exists for their relationship with Indo-European.

Early in the 2nd millenium B.C. two important developments took place in Anatolia: the arrival of Indo-European-speaking peoples and/or their relatives in a related branch (e.g. Hittites) from the north (as is suggested by cultural associations), and the introduction of literacy in the form of cuneiform writing diffusing from Mesopotamia. Cuneiform writing reached the Hittites via the Hurrians who were also invaders, but from northern Mesopotamia. (Hurrian and other non-Indo-European and non-Semitic languages of Mediterranean Europe and the Most Ancient East are listed under 2 in Languages of the World: Ibero-Caucasian and Pidgin Creole Fascicle One.)

Hittite (Hatti) was widely spoken throughout Asia Minor during the 2nd millenium B.C. The capital city of the Hittites was Hattusas (now Boğaz-köy) located in north-central Anatolia. The archives of the Hittite empire, found at Boğaz-köy, give ample evidence that there existed international communication in Hittite, Akkadian and Hurrian -- three languages belonging to three different language families, respectively: Indo-European, Semitic, and neither. The Old Akkadian cuneiform script which was used to write Hittite obscures the phonemic contrasts of Hittite, particularly by including redundant vowels not present in actual pronunciation. Hittite has only one stop series / p t k/ . Of particular interest in Hittite phonology is the presence of several laryngeal consonants, supporting the IE laryngeal hypothesis.

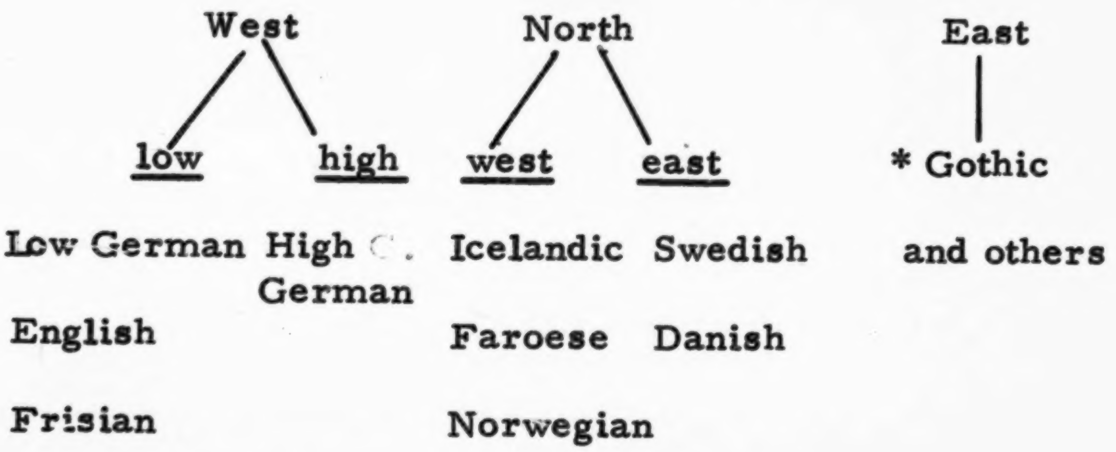
1.2. Some of the early splits from the parental Indo-European language left no descendants -- e.g. Tocharian A and Tocharian B, spoken in Central Asia and preserved in 6th century manuscripts found in Chinese Turkestan. What is often supposed to represent the earliest division of parental Indo-European is a division into eastern languages belonging to the satem-languages which have sibilants instead of velars in such forms as satem hundred (Balto-Slavic, Indo-Iranian, Armenian, and Albanian), and western languages which have velars in the cognates of the sibilant forms, as centum hundred (Germanic, Greek, Italic, and Celtic). Since Tocharian was located geographically among the eastern languages, it would be expected to be a satem-language, but this expectation is contrary to fact; Tocharian, spoken in Asia, turns out to be a centum-language, like Indo-European languages spoken in Europe.

This is cited as one of a half dozen overlapping features of typological samenesses among Indo-European languages that conflict with the family-tree model -- cited by Leonard Bloomfield (Language, N. Y. 1933), after Schrader; the other represent sameness which are:
shared by Balto-Slavic and Germanic (case-endings with [m] for [bh]);
shared by Celtic and Italic (passive-voice endings with [r]);
shared by Greek, Indo-Iranian and Armenian (prefix [é-] in past tenses);
shared by Italic and Greek (feminine nouns with masculine suffixes);
shared by Italic and Germanic (perfect tense used as a general past tense).

1.3. It is, as mentioned above, possible to chart the branches of the Indo-European family in terms of the daughter languages in each branch. Not counting the Anatolian branch of Indo-Hittite, whose descended languages have all become extinct, and instead viewing only the branches or sub-branches descended from Indo-European proper, it appears that one language or sub-branch in the Germanic branch became extinct (Gothic); a dozen or more languages in the Italic branch (or related in some way to Italic) became extinct; and two languages which may have once been members of the Armenian branch

(Phrygian and Thracian) also became extinct. There are other instances of language extinction in other branches of Indo-European, notably in Celtic (4. below). But the chief discussion of subrelationships within particular branches of Indo-European has probably been devoted to subrelationships among Germanic languages and among Italic languages.

The most generally accepted historical classification of the Germanic languages recognizes three groups: West Germanic, North Germanic and East Germanic:



Most of the difficulties in this scheme have to do with the position of West Germanic. Maurer believes that West Germanic must from its very earliest stage be separated into two nuclei of development, one in the Alps and upper Danube, the other in the lower Rhine and North Sea coast. According to this view then, instead of one West Germanic branch there were two. Other scholars would divide West Germanic into three original groups (Lower, Middle and Upper) and still others take West Germanic to represent an instance of historical convergence rather than divergence. The differences between English and German (Low or High) are greater than those between

German and the Scandinavian languages. Structurally English and German are today more remote from one another than are German and Icelandic; or again, lexically German stands closer to Danish and Swedish than it does to English.

The only languages in the Italic branch that have not become extinct are those descended from Latin--the so-called Romance languages (2, below). All others associated in one way or another with the Italic branch--more than a dozen--have become extinct:

Raetic (sometimes considered to be related to the non-Indo-European Etruscan language rather than to the Italic branch of Indo-European);

Venetic (shared some features of the Etruscan alphabet--e.g. vowels not preceded by consonants and consonants not followed by vowels or by the continuants /r n l/ are specified by a diacritic point on the relevant side of the letter; though Etruscan has not, Venetic has been demonstrated to be Indo-European, probably closer to Latin than to Illyrian but, in either case, in the Italic branch);

Lepontic (closely related to Ligurian);

Ligurian (an Italic or pre-Italic language which 'may well have been the earliest type of Indo-European spoken in Italy', according to Ernst Pulgram (The Tongues of Italy, Cambridge, 1958, p. 202); Ligurian was influenced by Celtic in northwestern Italy where the Ligurians were finally restricted in the country around the Gulf of Genoa);

Sicel (closely associated with Ligurian);

Oscan-Umbrian (two languages constituting the first of the three best known groups of languages in the Italic branchs);

Latino-Faliscan (two or more languages constituting the second of the best known Italic groups);

Sabellian (the third large group of Italic, including such names of dialects or languages as Paelignian, Marrucinian, Vestinian, Marsian, Aequian, and Sabine);

Sicilian;

Messapic (closely related to Illyrian; possibly an offshoot of Illyrian);

Illyrian (spoken in the Italic branch and hence a centum-language--but one spoken in the Albanian area in which Albanian is a satem-language; Illyrian continued to flourish in the Balkan peninnsula during the first few centuries of the Christian era).

1.4. **The family-tree model shows one main branch called Italic, including Latin, whose descendants are treated below as Romance languages (2). The next Indo-European offshoots treated below are the Germanic branch (3) and the Celtic branch (4). The family-tree model often represents Balto-Slavic as a single offshoot of Indo-European, though this represents an older view. We distinguish between a Slavic branch (5) and a Baltic branch (6). The next few offshoots of Indo-European treated below are represented today by one or two languages--Greek (7), Albanian (8), and Armenian (9); and though each of these is differentiated dialectically, Modern Greek is much more differentiated than the other two--in fact, there are two**

separate Modern Greek languages.

The remaining offshoot of Indo-European included in this report--but not in the sense of being the last to depart from a putative Indo-European homeland--is universally regarded as one main Indo-Iranian branch. Two sub-branches of this main branch are distinguished without controversy--Iranian and Indic. Though the Dardic languages unquestionably belong in the Indo-Iranian branch, postulation of its sub-relationship has shifted. First, Dardic was classified within the Iranian sub-branch (by Grierson, possibly because the nearest neighbors of Dardic are speakers of Iranian languages); next Dardic was reclassified within the Indic sub-branch--but also as a sub-coordinate Dardic sub-branch. Linguistic criteria--but different ones--are offered by different Indo-Iranian specialists for classifying Dardic as a member of the Indic sub-branch, and also for classifying Dardic as a sub-branch coordinate with Indic on the one hand, and Iranian on the other hand. For purposes of this report, we list the Iranian languages in one group below (10), without including Dardic among them, since Grierson did not cite common innovations to justify an Iranian-Dardic split from proto-Indo-Iranian. For different reasons, we list Indic languages in one group below (11), and then list Dardic languages finally (12). But this sectional segregation of Dardic, after Indic, does not reflect an independent examination and critical conclusion on whether common innovations justify sub-relating Dardic with Indic, or not.

All those who use this report on Indo-European languages will be happy to learn that relevant parts of the report were read critically by consultants who are specialists in that branch of Indo-European for which they supplied us with corrigenda and addenda incorporated herewith: Murray B. Emeneau of the University of California; Calvert Watkins of Harvard University; and Indiana University colleagues--Vladimir Honsa, Fred W. Householder, Harold L. Klagstad, Jr., Alo Raun, Albert Valdman, and Harry V. Velten.

ROMANCE DESCENDANTS OF LATIN IN THE ITALIC BRANCH OF INDO-EUROPEAN

2.0. Estimates of the number of Romance languages range from five (neglecting Rhaeto-Romance) to more than twice that number. This variability in counting separate languages reflects in part variability in criteria used and variability in time periods at which the estimates are made. Thus, there are five or six Romance languages, according to the criterion of one main national language for at least one politically autonomous country:

Catalan (spoken in Andorra, which is politically independent today)
French
Italian
Portuguese
Rumanian
Spanish

Until the 14th century, Sardinia was politically independent; by this and other criteria, including the use or preservation of a variety of Romance as literary language, there are eleven Romance languages; the list which follows is Ernst Pulgram's (longer lists than this, as in the Historical Atlas of the World, Rand McNally, 1961, are obtained by combining language names and dialect names):

1. Italian
2. Rumanian
3. French

4. Provençal

5. Spanish

6. Portuguese

7. Catalan

8. Sardinian

9.-11. Rhaeto-Romance languages.

The political status of Sardinia has been changing repeatedly since the 14th century. It was dominated politically by every national Romance country (except Rumania) until it finally became a part of Italy.

The literary status of each of the Rhaeto-Romance languages is questioned by one or another source cited by W. D. Elcock (The Romance Languages, London, 1960):

Ladin: written form dates from 18th century (Elcock); some attempts in lyrical poetry, a recent literary development (Bourciez); hardly any literature (Gartner).
Friulan (Friulian): least representative of the three Rhaeto-Romance languages, with written texts from 13th century (Elcock); important until the 16th century (Gröber); major poet, Piers Zorut, in 19th century followed by modern authors composing satirical comedies in prose (Bourciez); today humorous poetry is the main literary genre (Gartner).
Romansch: two standards, one for the Catholic and another for the Protestant population for whom there is a New Testament translation of the Bible (Elcock).

In counting separate Romance languages spoken today, it is necessary to recognize the possibility of language leveling. Dialect leveling is recognized universally: two different dialects through interaction of its speakers may become like each other. Two related languages, in similar circumstances, may become sufficiently like each other as to be reclassified as dialects rather than as separate languages. For example, it is true that Provençal and the kind of French spoken in and around Paris were once separate languages, according to the criterion by which lack of intelligibility (between speakers of different speech communities after some days or weeks of contact) serves to identify a separate language. According to that criterion, Provençal and the French of Paris and its environs have become dialects of a single language— after two world wars which brought Provençal and Parisian speakers together in one army, and after a generation of increased attendance at schools in which Standard French is taught; these and more complex factors are discussed by André Martinet (A Functional View of Language, Oxford, 1962). The emerging picture formulated by 'realists' includes socio-cultural factors in calculating whether there is or is not more than a single Romance language spoken in France today.

Socio-cultural factors may, however, not always be mutually confirmatory, and may cross-cut historical development (2.6, below). Though Provençal speakers may today be able to understand Parisian speakers after a shorter period of contact than it takes to learn a really separate language, the fact remains that the Provençal literary tradition is different than the Standard French literary tradition. If separate literary languages are counted, there are eleven such for Romance languages. If separate national languages are counted, there are a half dozen rather than a dozen. French is, to be sure, the national language of France; Spanish of Spain; Portuguese of Portugal; Italian of Italy; Rumanian of Rumania; and though there are three (rather than one) Rhaeto-Romance languages, only one of them, Romansch, is counted (since 1938) as an official language of Switzerland. All this reflects the political situation rather than the language situation which stands aloof from the political situation, but not entirely so. For example, there is some evidence that Catalan is linguistically less close to Spanish than it is to Provençal, even though Catalan speakers are politically Spanish rather than French—not counting those living in Andorra.

In order to approximate a more realistic picture than that obtained by listing a dozen literary languages or a half dozen national Romance languages, we enumerate separate languages in terms of Romance linguistic groups, with the groups presented in general geographic order, from west to east. According to this enumeration, there are three separate Ibero-Romance languages (2.1), and three separate Rhaeto-Romance languages (2.3, below); but there is today only one Gallo-Romance language which, accordingly, represents a group of dialects (2.2); so also, there is only one Italo-Romance language (2.4), and only one Rumanian language (2.5).

2.1. IBERO-ROMANCE

(1) The total number of Portuguese speakers exceeds 76 million.

Portuguese is the national language of Portugal, including its Atlantic island possessions, the Azores and Madeira; in this homeland, however, only

9,100,000 people speak Portuguese.

Portuguese is also the national language of Brazil, where it is spoken by 64,000,000 people. Another two million people in Galicia in northwest Spain and 200,000 people in the United States are native speakers of Portuguese. Portuguese is also said to be spoken by some 1,400,000 people in the Portuguese possessions in Africa—Guinea, Angola and Mozambique; by 250,000 in Portugal's former possessions in Asia—Goa, in south Asia; and by 4,000 in Macao, just south of Hong Kong; and by an unknown number of people in Portuguese Timor in the Lesser Sundas, the island chain east of Java.

Portuguese shows little dialect diversity; subdivision into dialects in general follows the provincial divisions of Portugal. It is possible, nevertheless, to distinguish the following four dialect areas, within which lesser differences occur in the regions listed after each:

Galician (Northern Portuguese), in Galicia in Spain, and in the northern Portugal provinces of Entre-Minho-e-Douro and Traz-oz-Montes.

Central, in Beira.

Southern, in Estremenho (including Lisbon), Alentejo, and Algarve.

Insular, in Madeira and the Azores; Brazilian Portuguese is most similar to this dialect.

The southern dialect, as spoken in Lisbon, is the basis for Standard Portuguese. Brazilian Portuguese, as just mentioned, is most like the Insular.

All dialects share the same phoneme inventory and differ only in phonetic actualization of some phonemes. The following phoneme inventory, given for Brazilian Portuguese by Robert A. Hall Jr., (SIL 1.15, 1943), could serve as the inventory for all Portuguese dialects.

p	t		k	i		u
b	d		g	e	ə	o
f	s	š		ɛ		ɔ
v	z	ž			a	
m	n	ñ				
	l	l̃				
	r	ř				

/ˊ/ stress

Hall analyzes [y] and [w] as nonsyllabic allophones of the syllabic vowels /i/ and /u/. Vowels are nasalized before a nasal phoneme, and the voiceless stops are palatalized before front vowels.

Dialect differences include the fact that the alveolar trill, /ř/ above, is phonetically a uvular trill in European Portuguese, but an alveolar trill in Brazil. The voiceless-voiced contrast among non-labial fricatives (sibilants) is neutralized (phonetically voiceless) in word-final position in European Portuguese. W. J. Entwistle (The Spanish Language) asserts that /ə/ is sometimes not a separate vowel phoneme, but a centering allophone of other vowel phonemes in unstressed syllables—at least in European Portuguese.

(2) Spanish, spoken by some 140 million people in various parts of the world, is the national language of Spain, including the Canary Islands. It is also the national language of all but one of the Latin American nations—the

nations of Central and South America (except Brazil), of the Guianas and British Honduras, as well as of Cuba, the Dominican Republic, and (with English) the (United States) Commonwealth of Puerto Rico. Spanish is also spoken as a native language in the United States—in California and the Southwest by long established Spanish Americans, as well as by recent immigrants from Mexico; some Spanish Americans came to these states directly from Spain, while others sojourned first in Mexico. The number of such old Spanish Americans is almost negligible when compared with the number of Spanish speakers in the East, particularly in New York, Chicago and Miami, including very recent immigrants from Puerto Rico and Cuba. Estimates of the numbers of native speakers of Spanish in particular countries, given in descending order of number of speakers are: Mexico 31,328,000; Spain 26,789,000; Argentina 15,264,000; Colombia 12,070,000; Chile 6,640,000; Cuba 6,466,000; Venezuela 6,030,000; Peru 5,090,000; Ecuador 3,500,000; Dominican Republic 3,033,000; El Salvador 2,600,000; United States 2,500,000; Puerto Rico 2,480,000; Uruguay 2,352,000; Honduras 1,900,000; Bolivia 1,750,000; Nicaragua 1,500,000; Guatemala 1,400,000; Panama 1,000,000; Paraguay 900,000; Costa Rica 754,000; Philippines 494,000; Africa 100,000. There are in addition some 500,000 Ladino speakers scattered in Turkey, the Balkans, Israel, and in the New World, particularly in New York City and Buenos Aires. Spanish is also spoken as a second language by most of the five million Catalan speakers, and by the 500,000 Basques in Spain, as well as by many Indians of Central and South America.

Castilian is used as a synonym for Spanish—especially by Latin Americans—to avoid identification with the former ruling power, Spain; it is also used in two senses as a dialect name—as the name of a dialect of Old Spanish and as the name of a Modern Spanish dialect.

In European Spanish at least five dialect areas can be distinguished.

Leonese and Asturian, in the north, are very similar to each other and preserve many earlier features lost in other dialects spoken in Spain.

Asturian shares with Aragonian the feature of the preservation of initial [f], but differs in such features as having [ye] instead of common Spanish [ī] before [l^y] (for example, [kastyel^{y}a] for Castilla).

Aragonian, in the east, is distinguished from the other dialects spoken in Spain by such features as [pl^y, kl^y] where other dialects have [l^y] (for example, [pl^{y}eno] for *lleno*, [kl^{y}ave] for *llave*); it shares some other features with the Castilian dialect, but also shares features with Catalan—a different language, in the sense of not being a Spanish dialect.

Castilian, in north central Spain with Madrid and Toledo as the principal centers, is the basis for Standard Spanish and as such has had influence on all other Spanish dialects. Other dialects in Spain have largely leveled with it, except in secluded rural areas. The subdialect of Toledo differs from that of Madrid in such features as having [y] instead of [l^y] (for example, [kastiya] for *Castilla*).

Ladino, the dialect of the Sephardic Jews whose forebears left Spain in the fifteenth century, has preserved many features of the Castilian dialect of that period.

Andalusian, in the south, is distinguished by such features as the loss of final consonants and the replacement of [s] by [x] or by zero before [k] or [t] (for example, [frexko] for *fresco*; [teta] for *testa*). Since the early emigrants came from southern Spain (largely through the port of Cadiz) to Latin America, the Spanish of Latin America, though further differentiated, retains many features of the Andalusian dialect.

The phonemic inventories of two dialects of the Spanish of Spain are now given, after V. Honsa (in press, Hispania, 1964). The first of the two lists the phonemes of Standard European Spanish; the second lists the phonemes of Andalusian Spanish.

p	t	č	k	i		u
b	d		g	e		o
f θ	s		x		a	
m	n	ñ		plus stress		
		l̃				
	l					
	r r̃					
w		y				

Voiced stops have both stop allophones and fricative allophones: /b/ [b], [β]; /d/ [d],[ð]; /g/ [g], [γ]. The stop allophones occur after interruption of the central oral air stream (i.e. word initial, after nasals, after lateral), and the fricative allophones occur elsewhere.

The phonemic status of [y] and [w] in Spanish has been discussed at length by Bowen and Stockwell and Saporta (Lg 31. 236-40, 1955, 32. 287-92, 1956), and works cited therein. Whether they are treated in a particular analysis as separate phonemes or as allophones of the vowels /i/ and /u/ depends on the analysis of stress, internal open juncture and syllable boundary.

Compare Standard European Spanish (above) with Andalusian:

p	t	č	k	
b	d		g	
f	s			h
m	n	ñ		
	l			
	r	r̃		

i		u
e		o
	a	

plus stress

Andalusian has two subdialects, the s(or Seseo) subdialect, and the θ (or θeθeo) subdialect. The subdialects share the same phonemic system, but differ in having /s/ as phonetically [s] in the s dialect and [θ] in the θ dialect.

The following phoneme inventory is that of Spanish as spoken in Mexico City (after H. V. King, SIL 10, 1952):

p t č k

b d ǯ g

f s š x

m n ñ

l

r r̃

w y

i u

e o

a

plus stress: strong /´/

weak /ˇ/

The phonemic inventory for Spanish spoken in Guayaquil, Ecuador, differs from that of Spanish spoken in Mexico City only in having $/l^y/$ rather than /ǯ/.

Colloquial Choco Spanish, spoken on the northern coast of Colombia as given by Jacob A. Loewen, (IJAL 26. 330-334, 1960), differs in phonemic inventory from Mexico City Spanish only in lacking /š/.

The phoneme inventory for Colorado Spanish as given by Wick R. Miller, (IJAL 25. 147-153, 1959), after Trager and Valdez (1937), differs from that for Mexico City in lacking /ǯ/.

(3) Catalan is spoken by five million people of whom 4,800,000 live in the Spanish provinces of Catalonia, Valencia and the Balearic Islands; the remaining 200,000 live in the Roussillon region of France (just north of Catalonia) and in Alghero on Sardinia.

Most speakers of Catalan also speak, or at least understand, Spanish, from which an increasing number of loan words are being borrowed into Catalan.

In phonological development from Proto-Romance, Catalan is archaic, having preserved—as has the Provençal dialect of French—final consonants (from Latin); and [u] which is also preserved in Spanish but not in Provençal.

Three dialects of Catalan are distinguishable:

Catalonian, the most uniform of the dialects—i. e. with little subdialect diversity.

Valencian, which is more diversified, through the influence of Castilian Spanish, and differs from Catalonian in that Catalonian [v] corresponds to Valencian [b] , intervocalic [z, ʒ] to [s, c], and [ž] to [č].

Insular Catalan, spoken on the Balearic Islands, is the only remaining Catalan dialect that is labelled as such.

The following inventory is that of (Valencian) Catalan (after V. Honsa, in press, Hispania, 1964):

p	t	č	k	i		u
b	d	ǯ	g	e		o
f	s	š		ɛ		ɔ
	z	ž			a	
m	n	ñ				
	l	l̃				
	r	r̃				

2. 2. GALLO-ROMANCE

(4) French is the native language of over 52 million people altogether; it is spoken as a second language by at least another 12 million people. Besides the 42 million native speakers of French in France and Corsica, French is the native language of 5 million Canadians concentrated mainly in the province

of Quebec; of over 4 million Walloons in Belgium, where it is one of the official languages of the country; and in Switzerland, French is spoken by over a million people in the five cantons east of the French border (Neuchâtel, Vaud, Genève, Valais and Fribourg). French is also the native language of the 20,400 citizens of Monaco; of 99,750 people in the Val d'Aossa in Italy; perhaps French is spoken by as many as 1,300,000 people in the United States of whom at least 100,000, largely in Maine and Louisiana, are descendants of Colonial French ancestors. Most of the 3 million people in France who are not native speakers still speak French as a second language. Some 5 million people in Africa—largely in Algeria, Tunisia and the former Belgian Congo—are said to speak French. Almost another 4 million people in Vietnam are said also to speak French as a second language. French is also spoken by at least one-third (25,000) of the people of French Polynesia, largely those having contact with the city of Papete, where Tahitian is the native language. So also, French is spoken by about ten per cent (400,000) of the people of Haiti, whose native language is Haitian Creole, quite definitely a separate language; so also, French is spoken by the educated people of the Creole speaking islands of the Lesser Antilles (particularly in Martinique, Guadeloupe, and St. Lucia).

The center of the Romance dialect which developed into Modern French was the Roman province of Gallia (Gaul), conquered by Caesar in the middle of the first century B. C.. The boundaries of this ancient province correspond roughly to the present boundaries of France.

Before its Latinization, France was occupied mainly by Celtic speakers, not to mention the Greeks who settled Massilia (modern Marseilles). There is some evidence that in Aquitania (the modern Gascogne) and on the Riviera still other languages were spoken—extinct languages called Iberian and Ligurian, respectively. The Celtic language of Gaul was completely replaced by a Romance language. (Breton, the modern Celtic language of Brittany, was introduced from Britain in the early Middle Ages.)

It is possible to distinguish five major dialects (with some further subdialect differentiation) for Modern French spoken in France and the adjacent regions in which French is spoken (paragraphs *2. 2. 1* to *2. 2. 5*, inclusive):

2. 2. 1. Standard French, which is essentially the dialect of the Ile de France, or more precisely of Paris, has had such a powerful influence on the surrounding dialects, particularly those to the south, that distinctions between Standard French and other dialects no longer constitute an insurmountable language barrier (A. Dauzat, L'Europe Linguistique, Paris, 1940; André Martinet, A Functional View of Language, Oxford, 1962). Distinctions between other dialects are also said to have been more or less leveled by the widespread use of Standard French; such dialects are not derived from Standard French—they are, rather, still distinguished from Standard French (cited in spelling orthography) on the basis of differences such as those exemplified in the following phonological features (after Dauzat):

(1) the substitution of [k] for [š] before [a] of other dialects (*vache* is pronounced [vak], *chanter* is [kãte]);

(2) the preservation of Latin stressed [a] (*aimer* or *aimé* is pronounced

[ama], *pelle* is [palo]);

(3) the zero reflexion of Latin [t] which is otherwise reflected by [d](Latin rotundus [redõn] for *rond*, [amado] for *aimé*);

(4) the preservation of [s] before [k, t, p]([eskuta] for *écouter* , [testo] for *tête*);

(5) the substitution of [h] for initial [f]([harjesto] for *fenêtre*).

2. 2. 2. Northern dialect group, including two sets of subdialects: (a) Norman, Picard, Walloon; (b) Lorraine, Franc-comtois, Bourguignon, Angevin, Gallot, Poitevin, Saintongeais, Berrichon, and Bourbonnais. The first set of Northern dialects, (a), is distinguished from the second set, (b), by having feature (1) as given above (*2. 1. 1*). The Northern dialects are set off from the other dialects listed below by not having feature (2), as given above.

2. 2. 3. Franco-Provençal dialect group, including Lyonnais, Dauphinois, Vandois, Neuchâtelois, Valaisien, and Savoyard. The Franco-Provençal dialect is distinguished from the other dialects by having feature (3), above, which it shares with the Northern dialect (but unlike the Northern dialect, it does have feature (2)). Franco-Provençal also shares (3) with one subdialect of the Provençal dialect (which differs from Franco-Provençal in having feature (4)).

2. 2. 4. Provençal dialect group, including Alpine Provençal, Auvergnat (within which Haut is distinguished from Bas), Languedocien (also distinguishing Haut from Bas), Caussenard, Limousin, and Provençal of southern Provence.

Bas Languedocien is distinguished from Haut Languedocien by having feature (1), a feature which it shares with the Caussenard and Provençal subdialects, as well as with the Northern dialect (from which the Provençal dialect differs in having feature (2)), and with the Gascony dialect (from which the Provençal dialect differs in not having feature (5)). Haut Auvergnat differs from Bas Auvergnat and the Limousin subdialect in having feature (4), a feature shared with the other six subdialects of the Provençal dialect as well as with the Gascon dialect (from which the Provençal dialect differs in not having feature (5)).

2.2.5. Gascon dialect, which differs from all the others in having feature (5).

2.2.6. Most French phoneticians and linguists agree on the following traditional phonemic inventory for Standard French (cf. Robert L. Politzer, Teaching French, Boston: 1959):

p	t		k		i	y	u
b	d		g		e	ø	o
f	s	š		ɛ:	ɛ	œ	ɔ
v	z	ž			a	ə	ɑ
m	n	ñ					
	l	r			ɛ̃		ɔ̃
w	ɥ	j			œ̃		ã

Robert A. Hall, Jr., Colloquial French Phonology (SIL 4.70-90, 1946) adds the additional phoneme /ʔ / and reinterprets /ɛ:/ as /:t/ in the variant of Standard French he labels 'Slow Colloquial.' Hall's /ʔ / is phonetically 'slight faucal constriction, renewed syllable onset and optional glottal stop'; in oratorical or emphatic speech, words spelled with the letter h often have [h] as a free variant of /ʔ/. Hall's /:t/ is phonetically a 'shortened' [t], distinguished from /t/ only as the final (or pre-final before /r/) phoneme of words in which it is preceded by /ɛ/; thus /:t/ represents an alternative analysis of vowel length, which Hall would treat as a phonemic feature of vowels only in 'archaic' style, 'spoken but little, except in formal uses such as oratory and declamation, and in singing.'

The vowel inventory for Fast Colloquial Standard differs from that of Slow Colloquial Standard by the merging of [e] with [ɛ], [ø] with [œ], [o] with [ɔ] and [a] with [ɑ].

The phonemic inventory given in R. Jakobson and J. Lotz, Notes on the French Phonemic Pattern (Word 5. 151-8, 1949), differs from Hall's consonant inventory in lacking the /:t/, which is Hall's alternative analysis for vowel length. Vowel length is treated by Jakobson and Lotz as a concommitant feature of tense saturated vowels when not in word final position. Their analysis unites the [ʔ] and [ə] in a single phoneme which combines the features of both and is opposed to all the other phonemes.

Their overall inventory for vowel phonemes may be stated as:

i	y	u	ê	ø̂	ô	ẽ	ø̃	õ
e	ø	o		â			ã	
	a							

However, some of the oppositions are neutralized in certain positions so that all of their vowel phonemes do not occur in all positions of the word, making necessary the use of additional symbols for transcribing vowels in positions in which an opposition has been neutralized. This neutralization, particularly of the distinction between the phonemes /e/ and /ê/, as well as the differences in their phonetic qualities of vowels in different positions, has caused difficulties in the phonemicization of French vowels.

George L. Trager (French Morphology: Verb Inflection, Lg 31.511-539, 1955) and Knut Togeby (Structure immanente de la langue française, Copenhagen, 1951) reduce the vowel inventory further by positing internal junctures and by the extensive use of /ə/ as a cover symbol:

i	y	u
e	œ	o
	a	

In normal transition the mid vowels and /a/ are [ɛ œ ɔ ɑ]; when followed by N' they yield the nasal vowels, and when followed by an internal juncture they are [e ø o ɑ].

The analysis of André Martinet and Albert Valdman avoid the problem of purely positional neutralization—neutralization which is correlated only with position in a sequence of phonemes—by setting up separate ('coexistent') vowel systems for each of the relevant positions (i.e., all positions in which different vowel phones are in opposition).

Thus Martinet sets up at least two vowel systems for phrase-final syllables—one for syllables of the shape consonant-vowel, and the other for syllables of the shape consonant-vowel-consonant. The front unrounded vowels of these two systems would differ in having

for final CV:	for final CVC:
i	i
ɛ ɛ:	e
a a:	ɛ
	ɑ

Both eliminate /ə / from the inventory—Martinet has recourse to it only to account for contrasts involving 'aspirate h' (le hêtre vs. l'être); Valdman (Applied Linguistics — French, Boston, 1961) upgrades it to the morphophonemic level, i. e. the morphophoneme /E/ is reduced to the archiphoneme Œ or phonemic zero; the former in turn yields /ø/ or /œ/.

Further coexistent vowel systems are distinguished by Valdman (in the Voegelin-Valdman review of Martinet's Functional View of Language, IJAL 29. 274-83, 1963) for differences in oppositions in particular positions in particular styles of speech (e. g. 'normal' versus 'formal'). Where different styles show different numbers of vowels in a particular position, vowel distinctions of the style with more vowels are said to be neutralized in the style with fewer vowels. Thus, Valdman has at least five coexistent vowel systems for educated speakers of Northern French; the number would be greater if, for example, one were to specify the final consonant of final CVC syllables, since, for example /œ/ and / ɔ/ but not /ø/ and /o/ occur before /r/. These five systems differ in their mid and low oral vowels as follows (symbols in capitals represent the 'neutralizations' mentioned above):

Final CV:

e ø o

ɛ

a ɑ

Final CVC, Formal:

ø o

ɛ ɛ: œ ɔ

a ɑ:

Final CVC, Normal

o

Œ

Ɛ A ɔ

Medial Syllables, Formal:

e ø o

ɛ œ ɔ

a ɑ

Medial Syllables, Normal:

E O Ö

A

The simplest alternative is to analyze the French vowel inventory in terms of seven basic vowels of which the four lowest can generate two additional series by the application of two sets of series generative components; in addition /:/ can also be applied to /e/ yielding /ɛ:/, as appears on the following chart.

$$\begin{array}{ccccccc} & & i & y & u & & \\ \pm /\!:\!/ & \Bigg[& e & \text{\oe} & o & \Bigg] & \pm /\sim/ \pm /\hat{}\,/ \\ & & & a & & & \end{array}$$

The component /^/ is to be interpreted as tensing and raising for all three mid vowels, as lengthening for the rounded vowels, and as backing ± lengthening for /a/. Styles, positional systems, and dialects differ by the number of components applied to the basic vowels as well as by the domain of the series generating components.

2.3. RHAETO-ROMANCE

Rhaeto- (Raeto-, Rheto-) Romance, often treated as a single, separate Romance language, is sometimes classified as a member of the Gallo-Romance subbranch, sometimes as a separate subbranch. Rhaeto-Romance is spoken by over 450,000 people in the Alpine regions of southeastern Switzerland and northern Italy. The literature speaks of Rhaeto-Romance as a group of dialects rather than as dialectally differentiated languages because no common history for groups of dialects can be shown by shared innovations; indeed, the few features shared by all the Rhaeto-Romance dialects are archaisms rather than common innovations. The speakers of Rhaeto-Romance dialects have never belonged to a single cultural or political unit, nor—more relevant here—do these dialects taken together reflect a common history.

The dialects have been grouped together partly on the basis of their being sufficiently different from most dialects of French, Italian, and Rumanian as to make them not readily classifiable as dialects of any of these national Romance languages. They differ in such features as: the retention of final /s/ (which results in differences in the morphology); the retention of initial consonant clusters with /l/ as the second member (which became palatized consonants without /l/ in Italian and Rumanian; thus, Rhaeto-Romance Engadine [kle:f, pla'ʒair] corresponds to Italian [kya:ve, pya'če:re]); the diphthongization of stressed vowels in open syllables ([e]>[ei, ai]); retention of lexical items closer to Latin (Rhaeto-Romance caput, Italian testa, French tête).

It has been argued that all of the Rhaeto-Romance dialects should be considered as conservative dialects of Italian—the western ones of the Lombard group and the eastern ones of the Venetian group. Until further study of the Rhaeto-Romance 'dialects' lends more support to this proposal, we will treat each of the 'dialects' as a separate language, and thereby obtain three languages—not merely because they are non-contiguous as speech communities but also because they do not form a natural linguistic group as dialects of one language would.

The three, geographically separated, Rhaeto-Romance languages are then:

(5) Romansch (Grishun, Rumauntsch) is spoken by some 48,000 people in the Swiss canton Grisons. Romansch enjoys greater prestige than the other two since it has been recognized as one of the official languages of Switzerland since 1938. Most speakers of Romansch are bilingual in German. The dialects of Romansch are:

Oberland Grishun, spoken in the Upper Rhine Valley, which has been particularly influenced by German;

Engadine, in the Inn Valley as far as the Austrian border, is divided into Upper and Lower Engadine dialects, the Upper one having been more influenced by Italian.

(6) Ladin, spoken by some 12,000 people in the Southern Tyrol (Alto Adije and the Dolomites) in Italy, has been influenced by Bavarian German. West of the Adige River Ladin is influenced by—or belongs to—Lombard dia-

lects of Italian; but east of the Adige Ladin is influenced by—or belongs to—Venetian dialects of Italian.

(7) Friulian, spoken by over 400,000 people in the Italian province of Udine, has also been influenced by—or is a subdialect of—Venetian Italian.

2.4. ITALO-ROMANCE

(8) Italian is the language of sub-alpine Italy, where it is spoken by some 49,000,000 people altogether. Of these, 20,000 live in the independent Republic of San Marino; 42,850,000 in the rest of Peninsular Italy; 4,700,000 in Sicily; and 1,400,000 in Sardinia. Italian is also spoken in Ticono canton and in three valleys of Grisons canton in Switzerland by 200,000 people; and in France, including Corsica, by 1,000,000 people; and in Yugoslavia by 300,000 people. In the New World, Italian is spoken in the United States (3,500,000), in Argentina (1,200,000), in Brazil (500,000), and in Canada (100,000). In Africa, Italian is spoken in Somalia (5,000), Ethiopia and Libya.

There are a large number of very divergent Italian dialects; in fact, Italian is more differentiated dialectically than any other national Romance language. The nature of the divergence of Italian dialects is such that although neighboring dialects are quite mutually intelligible, speakers of geographically distant dialects, as Venetian and Sicilian, would be surely unable to communicate unless they used some Standard Italian as a crutch for communicating, or unless they learned each other's dialects. Sardinian dialects are so very divergent from other Italian dialects that they are frequently classified as still constituting a language separate from Italian, as Sardinian appears to have been in some reconstructions (*2.6*, below).

Italian dialects have been variously classified as belonging to from two (Northern versus Central-Southern) to seven major groups, Northern, subdivisible into Venetian versus the rest (Gallo-Italian):

Venetian

Piedmontese

Ligurian (Genovan)

Lombard (Milanese), also spoken in a small area in southern Sicily

Emilian.

Central, subdivisible into Tuscan, including Corsican (or opposed to Corsican) versus the rest (then called Central or Latian):

Tuscan, subdivisible into Florentine, West Tuscan and South Tuscan

Corsican

Latian (including Latian proper, the subdialect of Rome)

Umbrian

Marchigiano,

Southern:

Abruzzese

Campanian (including Campanian proper or Neopolitan)

Apulian

Calabrian

Sicilian.

Sardinian:

Gallurese, in northern Sardinia Logodurese, in central Sardinian Campidanian, in southern Sardinia.

The Northern dialects are distinguished from the others by such features as: the lack of the phonemic distinction of long consonants, and the replacement of the short-long distinction by a fortis-lenis distinction (/s/ versus /z/); and by the replacement of [ky] and [gy], by [č] and [ȥ̌]. The Gallo-Italian subdialects distinguish rounded front vowels (y, œ) from unrounded front vowels, as does French.Ligurian /l/ and /r/ are dropped intervocalically. In Venetian, which is more like Tuscan than are the other Northern Italian dialects, intervocalic /t/ is frequently dropped; and /č/ and /ȥ̌/ of other dialects are replaced by /c/ and /ȥ/ or, sometimes, by /s/ and /z/.

The Tuscan dialect, or more specifically the subdialect of Florence, minus certain local peculiarities—as the replacement of initial [k] before [a] by [x] (for example, casa pronounced [xasa])—is the basis of Modern Standard Italian. Tuscan is more conservative (in respect to Latin) than the other peninsular dialects of Italian, but less so than Sardinian. A phonological feature peculiar to Tuscan is the replacement of intervocalic [r] of other dialects by [y].

The Southern dialects are distinguished from the others by such features as the replacement of initial [py] of other dialects by [ky] and of intervocalic [ll] by [dd].

The Central and Southern dialects share those features which distinguish Eastern Romance as opposed to Western Romance features of the Northern

Italian dialects. They retain intervocalic stops and unstressed final vowels.

The Sardinian dialects, often treated as a separate language, are treated in E. Bianchi, La Lingua Italiana (Florence, 1943), as part of the Central-Southern group of dialects because the northern Sardinian dialect (Gallurese) shows marked similarities to the dialect of southern Corsica and the southern Sardinian dialect (Campidanian) shows marked similarities to the dialect of Sicily. The central Sardinian dialect (Logodurese) differs most from other Italian dialects, having more conservative features, as the retention of final consonants in suffixes and of initial [k] and [g] before front vowels.

The following inventory of segmental phonemes for present-day Standard Italian is adapted from Robert A. Hall, Jr. (Descriptive Italian Grammar, New York, 1948):

p	t	c	č	k	i		u
b	d	ʒ	ǯ	g	e		o
f	s		š		ɛ		ɔ
v						a	
m	n		ñ				
	l		l^y				
	r						

All consonants occur with a series generating component of length, treated as geminate clusters by Hall. Word medially /š l^y n^y/ occur only doubled. Consonants occur only in word-initial and word-medial positions, except in loan words and poetry. All clusters of more than one vowel include /i/ or /u/

(phonetically [y] and [w]) as the second and/or third unstressed vowel. Vowels occur with four degrees of stress, the primary degree of which is often accompanied by non-phonemic vowel lengthening (i.e. stress is distinctive, and length is redundant). This redundant lengthening is not found among Northern Italian dialects. Among Northern Italian dialects, where length distinctions among consonants are not made, a phonemic distinction is made between /s/ and /z/.

2.5. RUMANIAN (BALKAN ROMANCE)

(9) Rumanian (Roumanian, Romanian) is the native language of over 19 million people. Of these, over 16 million live in Rumania; 2 1/2 million in the Soviet Union; 875,000 in Yugoslavia, Bulgaria, Greece and Albania; and 75,000 in the United States.

Four main dialects of Rumanian are distinguished:

Daco-Rumanian (Dako-Rumanian, Rumanian proper), spoken in Rumania, in Russia east of the Dnestr River, in Yugoslavia and Bulgaria, includes the following subdialects distinguished on the basis of some lexical differences:

Walachian or Southern Daco-Rumanian, consisting of the dialects of Mutenia and Oltenia, which, particularly as spoken by educated people in Bucharest, forms the basis of Standard Rumanian;

Transylvanian or Western Daco-Rumanian, including Banat, west of Transylvania;

Moldavian or Eastern Daco-Rumanian, spoken in Moldavia and the adjacent Ukraine, including a Bessarabian subdialect, the

Bukovina subdialect, spoken in the Ukraine north of Rumania, and perhaps the Dobrugea subdialect spoken on the right bank of the Danube in Rumania and northeastern Bulgaria.

Macedo-Rumanian (Arumanian, Aromanian), spoken in the Pindus Mountains of southern Yugoslavia, northern Greece, and western Albania.

Meglenitic (Megleno-Rumanian, Meglenite), spoken in the Meglen region northwest of Salonika in Greece (by 12,000 people in 1940).

Istro-Rumanian, spoken on the Istrian Peninsula of Yugoslavia (by 1,644 people in 1921).

Long geographical separation and contact with different languages has created sufficient differences between the four main dialects of Rumanian to make interintelligibility difficult. The greatest differences among the dialects are in vocabulary, each having borrowed words from different neighboring languages—Daco-Rumanian from Slavic languages (through church and administrative influence) and from Hungarian and German (there are some half million Hungarians and another half million Germans in Transylvania in Rumania); Meglenitic borrowed most heavily from Bulgarian; Istro-Rumanian from Italian and Serbo-Croatian; and Arumanian from Greek.

Though Daco-Rumanian has preserved more Latin words than the other three dialects, it has done so with more phonological innovations than the other three. Meglenitic is said to be linguistically intermediate (from the point of view of phonological changes) between Daco-Rumanian and Arumanian. Istro-Rumanian is said to be the most difficult to understand of the other dialects for speakers of Dako-Rumanian.

The phonemes of Standard Rumanian are discussed by Frederick B. Agard in three papers: Noun Morphology in Rumanian (Lg 29. 134-42, 1953); Review of Petrovici's Kann das Phonemsystem einer Sprache durch fremden Einfluss umgestaltet werden? (Lg 34. 297-303, 1958); Structural Sketch of Rumanian (Lg Monograph 26, 1958). Agard's inventory of the phonemes represents one possible alternative phonemicization:

p	t	c	č	k	
b	d		ǯ	g	
f		s	š		h
v		z	ž		
m	n				
	l				
	r				
w			y		

i	ɨ	u
e	ə	o
	a	

plus stress

This inventory implies the treatment of palatized and labialized consonants as clusters of consonant plus /y/ and consonant plus /w/, respectively.

The alternative analysis of Petrovici sets up a four-way contrast in single consonants: plain (as given in Agard's inventory) versus those with an additive component of palatalization, versus those with an additive component of labialization, versus those with both additive components. The inventories of other than the plain consonants are inherently limited by the non-occurrence of /w/ in the labialized series—that is, there is no /w^w/; and by the non-occurrence of the palatals /č ẓ̌ y/ in the palatalized series. These are homorganic limitations, and hence not fortuitous.

But, in addition, fortuitous limitations also occur in the series with additive components—e.g. only /l/ and /k/ occur with both the additive components of palatalization and labialization at the same time. Also, restrictions in distribution are much greater for the consonants with additive components than for those without—e.g. those with the additive component of labialization occur only before the stressed vowel /a/.

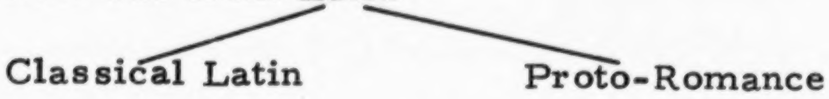
The inclusion of palatalized and labialized unit phonemes in the consonant inventory makes it possible to reduce the vowel inventory from a seven vowel systen (see above) to:

i	u		i
e	o	or even to	e
a			a

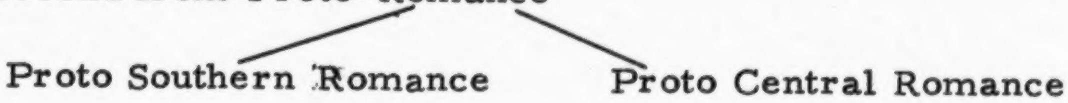
since the central vowels (see above) can be treated as the allophones of front vowels which occur after plain consonants and the back vowels can be treated as the allophones which occur after labialized consonants.

2.6. The historical development of the Romance branch from Proto-Indo-European is quite another problem, as mentioned above (1.2 and 1.3), than the differentiation of Latin into the daughter languages called Romance (2.1 - 2.5, above). In the post-Latin development of Romance, some languages split off from others only to become extinct before the 20th century (e.g. Dalmatian). According to Robert A. Hall, Jr., a series of successive splits are attestable in The Reconstruction of Proto-Romance (Lg 26.6-27, 1950):

The first was from Latin

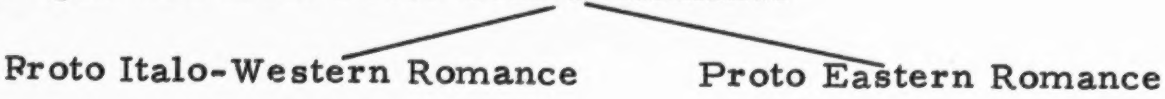


The second from Proto-Romance

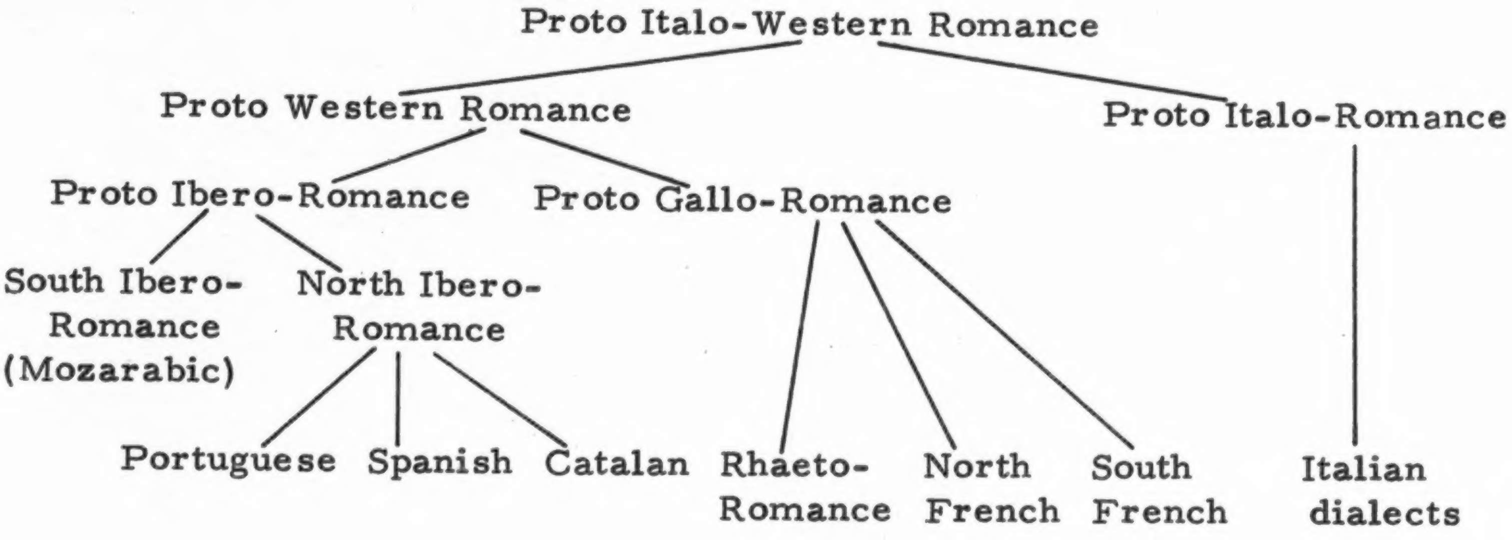


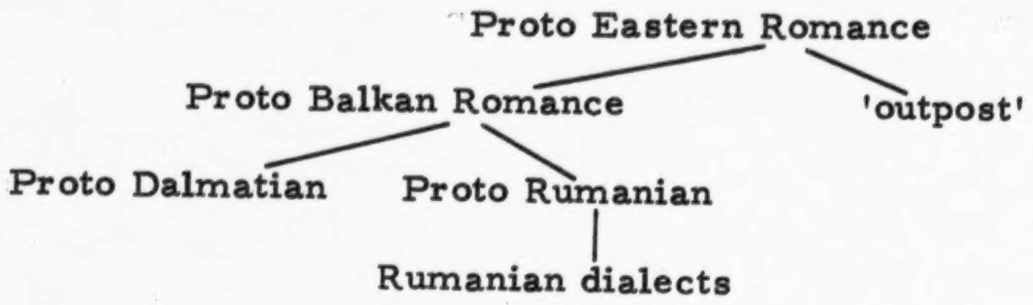
Since the daughter languages of Proto Southern Romance are listed as Sardinian, Lucanian, and Sicilian, the historical development cross-cuts the modern affiliation of dialects. Modern arguments for affiliating Sardinian and Sicilian with other more or less divergent Italian dialects have been alluded to above (2.4).

The third split was from Proto Central Romance



The remaining part of Hall's chart shows two parallel developments, on the western sides of Romance and on the eastern side of Romance.





Historical developments occasionally cross-cut modern dialect affiliations, as indicated above. Major divisions of languages and dialects are also reflected by modern isoglosses. Thus, all scholars agree that there is a boundary dividing the Romance language area (Romania) into two parts. It cuts across Italy from La Spezia on the Tyrrhenian Sea to Rimini on the Adriatic Sea. A bundle of isoglosses separates the Southeast and East (Eastern Romania) from the West (Western Romania).

GERMANIC

3. In spite of the various opinions as to the internal relationships of the Germanic languages, scholars have reached general concensus in reconstructing Proto-Germanic as making the following phonemic distinctions:

p t k   i: u: i u
b d g   e: o:
f θ x   a: a
s
z
m n
l
r
w y

Note: /i/ [i ~ e], /u/ [u ~ o], /b/ [b ~ ƀ], /d/ [d ~ ð], /g/ [g ~ γ], /x/ [x ~ h].

Some Germanic scholars include a sixth long vowel making the co-existent long vowel system completely symmetrical:

i: u:
e: o:
æ: a:

The differentiation of Proto-Germanic from Proto-Indo-European, which took place around 300 B.C., is known as the first Germanic soundshift, which was formulated by Grimm and later modified by Verner.

The 'Urheimat' of the Germanic languages is a subject of some dispute; however there seems to be general agreement that northern Germany along the Baltic, and what is now southern Scandinavia were the areas inhabited by Germanic tribes speaking a more or less mutually intelligible language.

Scholars of Germanic differ on the terms to be used for the regional differences that must have existed among speakers of Primitive Germanic. This is due in part to the lack of certainty concerning where, exactly, different forms were spoken, and how significant these differences were. Without a doubt, the incipient divisions of Primitive Germanic, while mutual intelligibility was still possible, whether practiced or not, formed the basis for the broad and diversified changes which took place later in the history of the Germanic language family.

The great Germanic tribal migrations of the early centuries of the Christian era parallel the linguistic diversification rather closely, so that English resulted from the invasion of the British Isles by the Angles, Saxons and Jutes; High German from the southward migration of the Alemanni, Bavarians and Lombards; East Germanic from the tribal dispersal of the Oder-Vistula group (Goths, Burgundians, Vandals; etc.); and West (Old) Norse from the westward seafaring of the Vikings. The groups that did the least amount of migrating, i.e. the East Scandinavians and some of the Saxons and Franks, linguistically came to constitute East Scandinavian and Low German (including Dutch-Flemish), respectively.

In terms of areas of high mutual intelligibility, there are only about five modern Germanic languages:

English	300 million speakers
Frisian	1/3 million speakers
Netherlandic-German (including Afrikaans and Yiddish)	115 million speakers
Insular Scandinavian	1/4 million speakers
Continental Scandinavian	18 million speakers.

A total of over two and a half times as many Germanic languages is obtained if one counts the languages in literary-national terms:

English	
Frisian	
Low German	Netherlandic-German
Dutch-Flemish	
Afrikaans	
Luxemburgian	
Standard High German	
Yiddish	
Icelandic	Insular Scandinavian
Faroese	
Nynorsk-Bokmål	Continental Scandinavian
Danish	
Swedish	

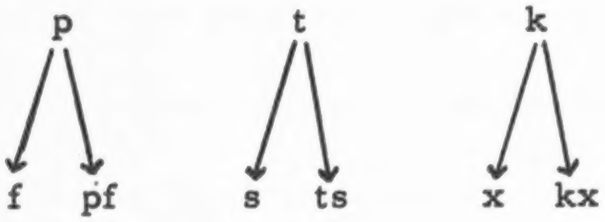
The five areas of mutual intelligibility mentioned above do not entirely reflect the historical development of the Germanic languages. For example, High German which split off from the rest of West Germanic by virtue of the second soundshift, is a member of the Netherlandic-German intelligibility area which includes Low German (in the broader sense) but not English or Frisian. Accordingly one might well speak of Insular West Germanic (English) and Continental West Germanic minus Frisian (Netherlandic-German).

A very similar situation exists for Scandinavian in which, in terms of gross mutual intelligibility, Norwegian belongs with Danish and Swedish, but historically it belongs with Icelandic and Faroese. Here one might speak of Insular North Germanic (Icelando-Faroese) as against Continental North Germanic (Norwegian, Swedish, Danish).

NETHERLANDIC-GERMAN

A western group has traditionally been recognized within the Germanic branch of Indo-European. This western group includes English, Frisian, Dutch-Flemish, Afrikaans, Low German, High German and Yiddish. The most general way in which these speech forms can be subgrouped is in terms of the so-called second soundshift (zweite Lautverschiebung) which took place in the Germanic speaking area of the Alpine region beginning about 550 A.D.. This soundshift resulted in the splitting off of High German (including Yiddish, which began to develop in the late 14th century) from the other speech forms mentioned above (all of which remained unaffected by this particular soundshift).

The second soundshift in its fullest form involves the replacement of the stops p t k by corresponding fricatives and affricates as follows:

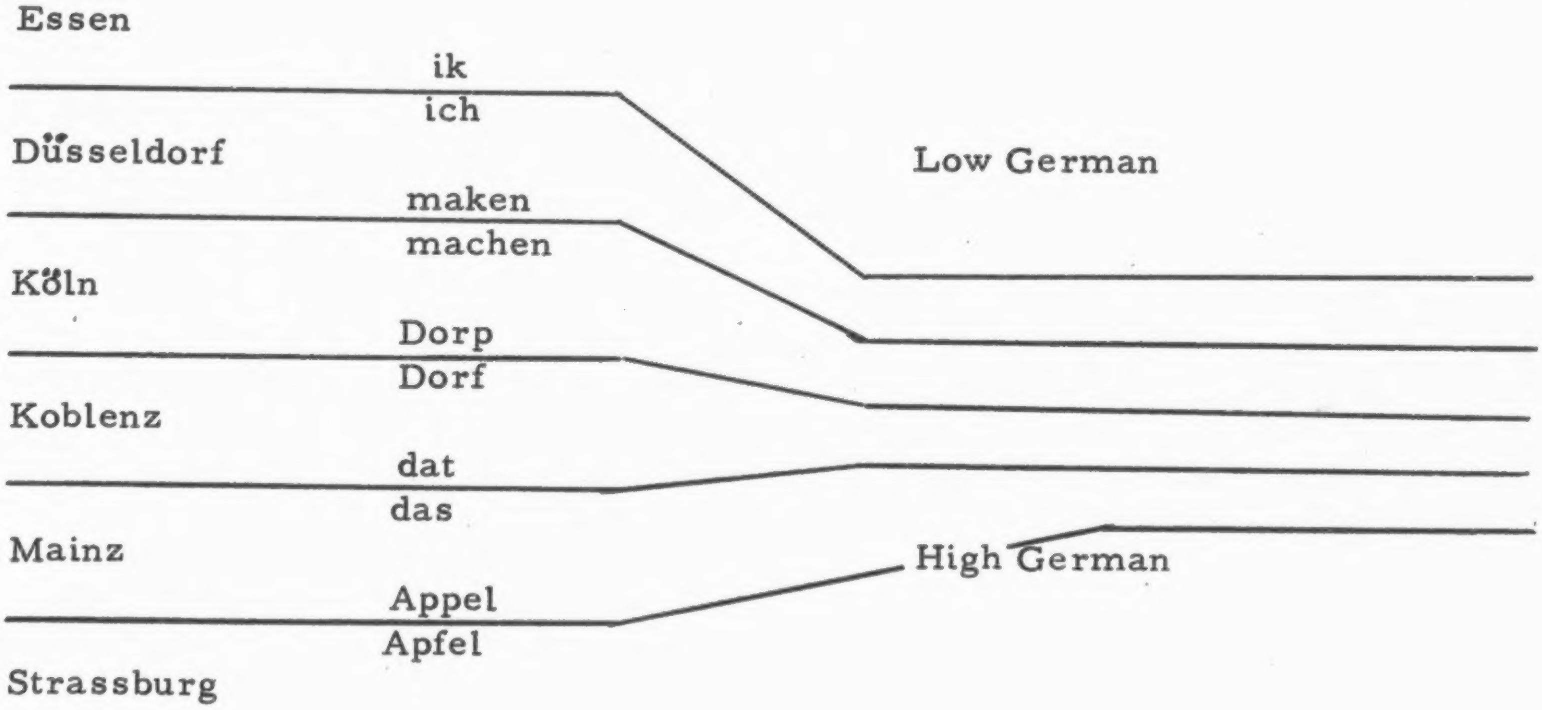


e.g. sleep vs. schlafen, apple vs. Apfel, water vs. Wasser, tongue vs. Zunge, make vs. machen, and in southern Alsace, Switzerland and southern Bavaria [kxint] or [xint] in contrast with the unshifted form Kind *child* occurring further north.

High German, which underwent this shift in varying degrees, is divided into three historical periods: Old High German, Middle High German and Modern High German.

Old High German (7th century to 1100) was fully inflected and underwent 'umlaut' with respect to the vowel /a/ only, e.g. gast *guest* with plural suffix -i became gesti. Middle High German (1100 to c. 1500) saw a great reduction of inflection coupled with the completion of the umlaut phenomenon which resulted in some cases in the introduction of front rounded vowels. Modern High German is not strictly a linguistic designation; it is rather a cover-term to account for developments occurring after High German emerged as a widely used literary form. The main impetus for this emergence was the Bible translation of 1522-34 by Martin Luther. Standard High German which quickly spread to the Low German speaking area in the north as well as the High German speaking area in the south, was itself based on the partly shifted (in terms of the second soundshift) Middle or Central German dialects.

The main bundle of isoglosses separating unshifted Low German (all of West Germanic except High German) from shifted High German runs across Germany from Düsseldorf on the Rhine (in the west) to Frankfurt on the Oder in the east, and continues on in a northeasterly direction up to Poland. This bundle of isoglosses is known as the Benrath line. In the west the various isoglosses fan out into what is known as the Rhenish fan, schematically as follows:



As far as standard literary languages go, the German speech area includes Dutch-Flemish, Afrikaans, Frisian, Luxemburgian, High German and Yiddish. Although Low German (in Germany mainly) has been written at various times and in a variety of conventions, it has never seriously competed with Standard High German in a literary sense. Even in the High German speech area where local dialects enjoy great prestige (e.g. Switzerland) Standard High German is the literary fashion.

From the point of view of spoken dialects and their degree of inter-intelligibility, it has been pointed out by many linguists that 'the entire Netherlandic-German territory from the North Sea to the Alps is a single dialect area with only gradual transitions from one village to the next' (Moulton, Encyclopaedia Britannica 16, 1964). The greatest cleavages are between 'Standard' (i.e. literary) and 'dialect' (i.e. colloquial). Bloomfield (Language, 1933) concurs with the view that the Netherlandic-German territory is a single language area in which neighboring dialects are mutually intelligible. Bloomfield, however, explicitly does not include Frisian in this cluster, whereas some linguists do. It appears that Afrikaans can also be included as a discontinuous part of the large Netherlandic-German speech area since intelligibility tests between Dutch and Afrikaans indicate that Afrikaans speakers understand approximately 95% of Standard Dutch, and Dutch speakers understand about 76% of Afrikaans (the latter figure is lower largely because Afrikaans contain a good number of loans not familiar to Dutch speakers). For Yiddish we have no specific information as to its degree of intelligibility with the High German dialects of the middle Rhine on which it is based (see below).

The Netherlandic- German speech area is generally divided into two main groups according to the second soundshift, with Low Franconian (Dutch Flemish) and Low Saxon (Low German) to the north (Anglo-Frisian, two separate languages, also belong with this group historically), and High German to the south. The High German area is further divided into Middle and Upper German, the former having been less affected by the second soundshift than the latter.

The following list of the Netherlandic-German speech communities is based mainly on Priebsch.

LOW GERMAN DIALECTS

By Low German we here mean to include Dutch, Flemish, Afrikaans and Plattdeutsch.

Dutch is a standardized development of Low Franconian with influence from Saxon and Frisian.

Flemish is also a standardized form of Low Franconian, differing only slightly from Dutch in spelling convention, as well as in its distinction of three genders, and the presence of a good number of French loans.

Afrikaans, one of the official languages of the Republic of South Africa (beside English) since 1925, is extremely similar to Dutch. However, Afrikaans has undergone a great deal of morphological leveling in comparison to Dutch. For example, there is no gender distinction among nouns, there is only one form of the definite article /di/, and both verbal and nominal forms have undergone considerable analogical change in the direction of 'regular' rather than 'irregular' forms. Verbs are marked for number only, not person. All tenses are formed with modals. The use of a double negative is an Afrikaans innovation.

PLATTDEUTSCH OR MODERN LOW GERMAN DIALECTS

Low Saxon dialects are spoken west of the Elbe, extending into Schleswig-Holstein. They have the first and third person plural indicative levelled to the second person in -t, e.g. wi, ju, sei makt. Over most of the Low Saxon area, the dative of the personal pronouns is mi and di with accusative either the same or mik or dik; only in the southeast between the Weser and the Elbe the dative is merged with the accusative as mik (mek), dik (dek). This southeastern area has as its nucleus the Eastfalian districts. Low Saxon dialects include:

Westphalian, spoken from the Zuyder Zee to the region of the Ems, including Münster and Osnabrück; characteristic pronunciation S-chinken;

Engrian, spoken in and and around Paderborn, Lippe, Göttingen, Grubenhagen, Hanover, Ravensburg;

Eastphalian, spoken in Hildesheim and Goslar, east of the watershed of the Werra and Fulda;

North Low Saxon (Ostfriesisch Platt), spoken between the Ems and the Weser, including the dialects of Oldenburg, Bremen, Lüneberg, and the Nordalbingisch dialects of Dithmarschen, Holstein and Lauenberg.

East Low German dialects include those of Mecklenburg and Pomerania as far as the Oder; those of the Priegnitz, the Altmark, Brandenburg, and the Ukermark, with a tongue of Central German protruding to Berlin and Potsdam from the south; dialects between the Oder and the Vistula which drop -n in inflections other than the gerund; West Prussian and unshifted East Prussian from the Vistula to the Lithuanian frontier and extending into the Low German settlements in the Baltic states in so far as these have survived. Prussian drops -n in all inflections.

Low Franconian (Niederfränkisch) is separated from Low German proper by a line passing west of Olpe, then between Barmen and Schwelm, Mülheim and Essen, Wesel and Dorsten and continuing northwards to the Zuyder Zee. Low

Franconian dialects include:

Bergish, spoken in Remscheid, Elberfeld, Solingen, Werden, and Mülheim an der Ruhr;

Limburgish, spoken between Düsseldorf in the south and Krefeld and Venlo in the north;

Geldersch, which is characterized by the High German shift -k>-ch in ich, mich, sich, auch, and the suffix -lich; the line limiting this shift is the so-called Ürdingen line beginning at Tirlemont, passing through Venlo and Cleves, going up the Rhine to Wesel and Duisberg, then south-east between Werden and Velbert, Elberfeld and Ronsdorf, Luttringhausen and Remscheid.

In some varieties of Low Saxon -en or -e mark the first and third present plural form of the verb; in other varieties -en or -e occur for all persons in the present plural.

Low German, or more specifically Low Saxon, in northern Germany, is rapidly being replaced by Standard High German. Before Luther's Bible translation, Low Saxon as the language of the Hanseatic League enjoyed great prestige, even as a literary language. But the Reformation and the cultural upheavals and displacements of two world wars in the present century have been crucial in the rapid decline of Low German. The same could be said for Frisian, which partly because of its great dialectal diversity, is losing ground in favor of Low German as well as High German. The situation for Dutch (or Low Franconian), strengthened by political separation, is quite the opposite of that for Low Saxon and Frisian.

MIDDLE OR CENTRAL GERMAN DIALECTS

These dialects form a broad belt whose northern limit is the Benrath line.

They include Luxemburg in the extreme west and southwestern Poland in the extreme east.

West Central German:

Middle Franconian, with it, dit, dat, etc., and with spirants v and f in geven, wif, including:

Ripuarian, spoken in Cologne, Aachen, keeps unshifted rp, rd, in werpen and hard, and -p in ūp;

Moselle Franconian, extending north to the Eifel, westwards into Luxembourg, eastwards to the Westerwald and Siegerland, the chief center being Trier, which shifts -rp->-rf-, -rd->-rt- and has uf;

Rhenish Franconian, extending northwards to Kassel, westwards into Lorraine (apart from the Moselle Franconian area), and southward to the northern part of Baden and Württemberg. The chief Rhenish Franconian towns are Frankfurt, Mainz, Darmstadt, Heidelberg, Giessen and Fulda. The Rhenish Franconian dialects are in some parts characterized by the elision of -n in the infinitive, e.g. esse, and the past participle (in agreement with Moselle Franconian of the Saar-Nahe-Moselle area). Rhenish Franconian dialects include:

Lorraine and Palatinate dialects extending to the Rheingau and the Odenwald from which regions the 'Swabians' of the Banat in southern Hungary migrated;

Hessisch-Nassauisch, with leib and gout for lieb and gut;

Low Hessian, extending to the Low German line.

East Central German:

Thuringian, with Rennstieg of the Thuringian Forest as the southern boundary and the foothills of the Harz as the northern boundary, has initial f- from older pf- from p- , but in the southern part has mpf; monophthongizes in the eastern part ei to ē and au to ō; shows a for old ĕ and elides in the infinitive, e.g. asse;

Upper Saxon, spoken in the former kingdom of Saxony and parts of the provinces of Saxony and Anhalt, lacks the distinction between b and p, d and t, g and k; monophthongizes ei and au like Thuringian;

Silesian, spoken in Prussian and Austrian Silesia, includes the dialects of northeastern Bohemia and most of German Moravia (Sudetendeutsch), viz. part of the German minority in Czechoslovakia, and of the Zips in Hungary. Silesian keeps the distinction between t and d, etc., and has f- for pf-, ŭ for ŏ, e.g. gegussa, and in the southern parts uses the Upper German diminutive ending -el, e.g. Liedl; the mountain dialects show -a for unstressed -en, e.g. beissa, gegussa;

Ostersgebirgisch, spoken in Bohemia on both sides of the Erzgebirge;

East Prussian, i.e. the dialect of the land of the 'Preussen' conquered and colonized by the 'Deutsche Orden' in the thirteenth century, was superseded by a Low German (East Elbian) dialect after 1600.

Standard High German is based on the Middle or Central High German dialects. Yiddish also is basically a Middle German dialect (from the 14th century) with heavy influence from Romance, Hebrew and Aramaic. The Romance element in Yiddish has been retained to some extent, even though the Romance speaking Jews of the 10th century shifted to the local German speech when they settled in the middle Rhine basin. Beginning with the Crusades these European Jews were scattered all over Europe until in the 14th century

Eastern Europe became their focal area. In this Slavic area, Yiddish adopted a good number of Slavic elements. Yiddish became the lingua franca of Jewry in the 19th century when the Jewish people migrated to all parts of the world. At the present time, Yiddish dialects reflect the European settlement pattern of the 17th century: Western (Germany, now extinct), Central (Poland) and Eastern (east of the Vistula). Standard literary Yiddish does not reflect a distinct dialect, but follows the conventions of the 19th century classical writers.

The local dialect of Luxemburg (which belongs to the Moselle subgroup of Middle Franconian and is locally known as Letzeburgisch) is recognized as an official language beside Standard High German and French, and through use in radio and education is also becoming a literary language. Letzeburgisch differs markedly from Standard High German; for example, the masculine singular form of the third person subject is hen, Standard High German er, English he.

UPPER HIGH GERMAN DIALECTS

The Upper German dialects are spoken in southern Germany, Austria, Switzerland and adjacent areas.

High or Upper Franconian includes South Franconian and East Franconian, and is spoken in the valley of the Neckar and the Murg, and of the Main from Würzburg to Bamberg, including the towns of Karlsruhe, Heilbronn, and Nuremberg and the Vogtland. The High Franconian dialects agree with the other Upper German dialects in shifting p and d; the western part of East Franconian agrees with Rhenish Franconian and Thuringian in eliding -n in unstressed syllables. In Henneberg on the upper Werra ks has become s, e.g. [hesse] Hexe, and w-in wer, was, wie has become b-.

The Alemannic group includes Swabian (spoken in Württemberg, northwest Tyrol to the Lech) which diphthongizes in Zeit and Haus and elides n before s with nasalization of the preceding vowel, e.g. [gā̃s] Gans, pl. [gē̃s] Gänse. Alemannic is divided into High Alemannic and Low Alemannic.

High Alemannic (including Schwyzerdütsch and the dialects of the southern part of the Black Forest and Vorarlberg) shifts k- to the spirant ch-, e.g. Chind, and does not diphthongize;

Low Alemannic (spoken in Alsace, apart from the northern strip, Baden south of Rastatt and north of the Feldberg, the town of Basel and part of Vorarlberg), has k- unshifted and in Alsatian changes ū to ü, e.g. [hüs], and diphthongizes only in hiatus and in final position, e.g. Reue, frei.

The Bavarian Austrian group includes the dialects of Old Bavaria with a southern set of dialects spoken in the Bavarian Alps, the Tyrol, Styria, together with the Heanzian dialect of the Burgenland, Carinthia, the Sette and Tredici Communi south of Trient in Northern Italy, and the enclave of Gottschee in Yugoslavia. Bavarian has a rounded form of [a] (as have many other dialects); long [ā] is very open. In the southern set of dialects k has become an affricate [kx] or aspirated [kh]. The following dialects are included in this group:

Central Bavarian spoken north of the Alps and in Upper and Lower Austria and Salzburg;

North Bavarian-Upper Palatinate spoken north of Regensburg, extending to

Nuremberg on one side, and western Bohemia on the other.

One of the interesting features of the Swabian dialect is the nasalization of vowels as in the verbs gãu, stãu, hãu, and lãu meaning *go*, *stand*, *have* and *let*, respectively (in Standard High German gehen, stehen, haben and lassen).

Upper German dialects as a whole can be distinguished from Middle German dialects in that they form noun diminutives in l whereas Middle German dialects have -chen or -ken.

High Alemannian enjoys great prestige in Switzerland, being used (besides the literary Standard High German) by all social classes in everyday communication. The larger cities each have their own well defined dialect which is used in its surrounding area.

A detailed study of German dialects was begun by George Wencker in 1876. Using 40 Standard German test sentences Wencker distinguished over 40,000 local dialects. Publication of this material began in 1927. It served as the basis of the on-going Deutscher Sprachatlas (in press).

The total number of speakers of Netherlandic-German is over 115 million: German (Low and High including Yiddish) 95 million; Dutch-Flemish (including Afrikaans) 20 million. In terms of political areas, these figures may be further broken down as follows:

Germany	70 million
Netherlands	12 million
Austria	7 million
United States	7 million
Belgium	5 million

Switzerland	4 million
South Africa	4 million
France	1.7 million
Poland	1.4 million.

In addition to the countries above, German, Dutch and Yiddish are spoken by less than a million speakers in each of dozens of countries around the world, particularly in Latin America.

NETHERLANDIC-GERMAN PHONEMIC SYSTEMS

The phonemic system of Standard High German (after William G. Moulton, Syllable Nuclei and Final Consonant Clusters in German, For Roman Jakobson, 1956) is:

p	t		k					
b	d		g		i	ü		u
f	s	š	x	h	e	ö		o
v	z	ž					a	
m	n		ŋ					
	l		r					
		y						

All of the vowels above also occur long; an additional vowel phoneme, /ə/, occurs only short and unstressed.

In some western High German dialects, and in Bühnenaussprache, there is an additional long low front vowel /ɛ/. Initial /č/ and /x/ (velar fricative) occur in loans; /ž/ occurs primarily in French loans.

The phonology of a Low German dialect which has developed in a fair amount of isolation in the Ukraine for well over a century before its speakers migrated to the New World, contrasts palatal consonants with velars, and lacks front rounded vowels altogether (Eric Mierau, Ph. D. dissertation, Indiana University, 1964):

p	t		k̭	k	
b	d		g̭	g	
f	s	š	x̭	x	h
v	z	ž			
m	n		ṋ	ŋ	
	l				
	r				

i	ə	u
e	a	o

Long vowels are treated as clusters of identical vowels, /ii, ee, oo, uu/; there is no contrast between long and short /a/. /ə/ represents the unstressed neutralization of /i e o u/. The phoneme /g̭/ has both palatal stop and palatal fricative allophones.

The phonemes of Dutch as given by Moulton (The Vowels of Dutch . . ., Lingua, 1962) are:

p	t	k	
b	d		
f	s	x	h
v	z	ɣ	
m	n	ŋ	
	l		

i	ü		u
e	ö	ə	o
ɛ			ɔ
		a	

r
w y

e: ö: o:
ɛ: ɔ̈: ɔ:
a:

Long /ɛ:, ɔ̈:, ɔ:/ occur only in loanwords..

The system given by Cohen, Ebeling, Eringa, Fokkema and van Holk (Fonologie van het Nederlands . . . , The Hague, 1959) differs from Moulton's by the inclusion of /g/ and by a slightly different treatment of the vowels:

Tense Vowels:	Lax Vowels:
i ü u	I ʊ
e ö ə o	ɛ ɔ̈ ɔ
a	ɑ

Vowels occuring only in loanwords and placenames include /i:, u:, ɛ:, ɔ:, ɔ̈:, ã/.

The phonemes of Afrikaans (as given in James L. Wilson, Ph. D. dissertation, Indiana University, 1964), are:

p t k
b d
f s x h
m n ŋ
l
r
y

i ü u
e ö ə o
ɛ ɔ̈ ɔ
ɑ

The contrast between long and short vowel occurs only with /ɑ/.

ENGLISH

English is today spoken by at least one tenth of the world's population, a total of roughly 300 million people. Of this number, about 250 million are native speakers and the remaining 50 million use English mainly as a foreign language. These figures do not include the numerous speakers of English as a second language in continental Europe, China, Japan, Southeast Asia, or Central and South America for which no estimates of the numbers are readily available.

The following list gives the political areas in which English is spoken by a million or more speakers (in descending order, in millions, asterisk indicating that English is used mainly as a second language).

United States	179
British Isles (including Eire)	55
Canada	13
Australia	11
Philippines	*11
India	* 7
South Africa	* 6
New Zealand	2.5
Jamaica	1.5
Tanganyika	*1.5
Kenya	*1.5

Uganda	* 1
Nigeria	* 1
Ghana	* 1

The high number of speakers of English as a second language in some parts of the world is due to the fact that English is the medium of instruction in the school systems of these areas. For example, English had this function in the Philippines after 1901 when 1,000 American teachers went to that country (Tagalog is now coming to be widely used as a medium of education). In the Republic of South Africa, most schooling is either in English or Afrikaans (even though the white population is only about 3 million—out of a total of 15 million), and children are sent either to an English school or an Afrikaans school.

Breaking down the number of speakers of English by gross areas results in the following picture:

New World	196 million
Europe	55 million
Pacific	23 million
Africa	15 million
Asia	8 million
Total	297 million

The history of the English language is generally taken as beginning with the invasion of the British Isles by the Angles, Saxons and Jutes in the fifth century of this era. These northern conquerors forced the native Celts to withdraw into marginal areas and proceeded to establish themselves both culturally and linguistically.

The dialects which these invaders spoke are collectively known as Old English (Northumbrian and Mercian Anglian, West Saxon and Kentish). During the latter part of the Old English period (c.450-1150) the language was substantially influenced by Norse speaking settlers primarily in the northeast. Until this time English had been relatively free of foreign elements.

Soon after the Norman conquest of England (1066) the Scandinavian influence on the language was cut short and replaced by a far more extensive Romance (French) influence, an influence that has continued to this day. The English of the period following the Norman conquest is known as Middle English (1150-1500).

Toward the end of the Middle English period and throughout the Modern English period (from 1500 onward) another extensive influx of foreignisms took place — this time mainly from Latin and Greek and to a great extent directly attributable to the impact of the Renaissance. The 14th, 16th and 19th centuries saw the largest amount of borrowing taking place. This borrowing included not only many classical terms but also elements like the prefixes ex-, anti-, co-, de-, inter- and re-.

The most influential of all the dialects of English is that of London. Although the original dialect of the London area was essentially of the West Saxon variety, the London English of the Modern period is of the Mercian (East Midland) variety. The southeastern variety of educated British English has come to be known by the term Received Pronunciation, or Standard British English, on which are based many non-European varieties of English. The great popularity of this dialect is due at least in part to the impetus it receives from both Oxford and Cambridge, a situation which, according to Eugen Dieth (A New Survey of English Dialects, Transactions of the Philological Society 32 (1946), has greatly inhibited the study of (other) English dialects.

The serious study of the dialects of English (both European and other) has not yet progressed to the point where clear statements of differences in various parts of their linguistic structures can be made. The general situation for English as a whole is that it is dialectally remarkably homogeneous in view of its vast extent geographically and its tremendously large number of speakers of various cultural and social backgrounds.

The most diverse dialects are to be found in the British Isles, particularly in some of the rural areas of Scotland. Were one to compare these dialects with, say, those of the highland areas of the Eastern United States, the degree of mutual intelligibility might reasonably be expected to be extremely low. But a comparison of, say, the educated Londoner's dialect with that of Boston would reveal only slight differences, and mutual intelligibility might well be near 100%. All dialects of English are regarded as mutually intelligible by such dialectologists as G. L. Brook (English Dialects, Oxford, 1963).

The following list of the English dialects of the British Isles is taken from the Scottish National Dictionary (for Scotland) and from W. W. Skeat (English Dialects ... . Cambridge, 1912) for England and Ireland.

SCOTS DIALECTS

Northern:

- Shetland and Orkney (insular);
- Cromarty, Eastern Ross and Black Isle;
- Caithness and Sutherland;
- Lower Banff, Aberdeen, Buchan and Deeside;
- Inverness, Nairn, Moray and Upper Banff;
- Mearns (Kincardine) and E. Angus (E. Forfar).

Mid:

- W. Angus (W. Forfar), E. and S. Perthshire, Stirling, Fife, Kinross, and Klackmannan;
- Linlithgow (W. Lothian), Edinburgh (Mid Lothian), Haddington (E. Lothian), Berwich and Peebles;
- E. and W. Dumbarton, S. Argyll, Bute, Renfrew, Glasgow, Lanark and N. Ayr;
- S. Ayr, Galloway and W. Dumfries.

Southern:

- Roxburgh, Selkirk and E. and Mid Dumfries.

IRISH ENGLISH DIALECTS

Ulster Scots (including Northern Ireland and the northern part of the Republic of Ireland);

Dublin (east-central Ireland);

Wexford (southeastern Ireland).

Some structural characteristics of Irish English are as listed below (after Brook, 1963); features (1), (2) and (6) are also characteristic of Scots dialects, and (4) occurs in some Scots dialects.

(1) w plus a where British English dialects have w plus ɔ.

(2) Middle English ē is generally retained in words like leaf, but British influence is causing some shifting to ī.

(3) θ is usually merged with t, as is ð with d, probably because Gaelic t and d are dentals.

(4) Final consonants are frequently unvoiced.

(5) There is no 'dark l'.

(6) The consonant r remains in all positions.

(7) Some tenses have unique constructions, e.g. the perfect using *after* and the continuous present using *do be* or *be:* I'm after eating my dinner. *I've had my dinner*. He do be cutting corn everyday. *He's cutting corn every day*.

ENGLISH DIALECTS

Northern:

Northumberland and N. Durham;

S. Durham, most of Cumberland, Westmoreland, N. Lancashire and hilly parts of W. Yorkshire;

N. and E. Yorkshire.

Midland:

Lincolnshire;

S. E. Lancashire, N. E. Cheshire, N. W. Derby;

S. W.. Lancashire, south of the Ribble;

Mid Lancashire, Isle of Man;

S. Yorkshire;

Most of Cheshire, N. Staffordshire;

Most of Derby;

Nottingham;

Flint and Denbigh;

E. Shropshire, S. Staffordshire, most of Warwickshire, S. Derby, Leicestershire.

Eastern:

Cambridge, Rutland, N. E. Northampton;

Most of Essex and Hertford, Huntingdon, Bedford, Mid Northampton;

Norfolk and Suffolk;

Most of Buckingham;

Middlesex, S. E. Buckingham, S. Hertford and S. W. Essex.

Western:

W. and S. Shropshire;

Hereford (except E.), Radnor, E. Brecknock.

Southern:

Parts of Pembroke and Glamorgan;

Wiltshire, Dorset, N. and E. Somerset, most of Gloucester and S. W. Devon;

Most of Hampshire, Isle of Wight, most of Berkshire, S. Surrey and W. Sussex;

N. Gloucester, E. Hereford, Worcester, S. Warwick, N. Oxford and S. W. Northampton;
Most of Oxford;
N. Surrey and N. W. Kent;
Most of Kent and E. Sussex;
W. Somerset and N. E. Devon;
Most of Devon and E. Cornwall;
W. Cornwall.

The dialects of American English are much less diversified than are those of Great Britain. For this one might give various reasons: their history is much shorter, and they have been relatively little isolated from one another when compared to the rural dialects of Great Britain.

Four major dialect areas have been recognized for American English:
Eastern (New England or northeastern United States)
Midland (east-central United States)
Southern (southeastern and Caribbean seaboard)
Northern and Western (Great Lakes region and most of the area west of the Mississippi).

One particularly significant feature of both Eastern and Southern American English is the lack of postvocalic / r/ , a feature often attributed to extensive British contact via important ports like Boston, New York, Richmond and Charleston. All told, some 40 million Americans speak 'r-less' dialects. The so-called intrusive r, occurring at word boundaries where one word ends in a vowel and the following word begins in a vowel, is common in Eastern American

English. Southern American English does not have this feature, nor does the dialect of New York City.

Specific examples of dialect differences in the United States are given below:

(1) Northern and Western /s/ corresponds to /z/ in other dialects in the word greasy.

(2) Eastern dialects have a low back vowel in words like path and half where other dialects generally have a low front vowel.

(3) The New York City dialect has a diphthong [ɨ^{i}] in words like bird and earth.

(4) A rounded vowel in the Midland dialects in words like fog and hog corresponds to an unrounded vowel in Northern and Western dialects.

(5) The palatal onglide before stressed vowels in words like Tuesday, dew and new occurs consistently in Southern, sporadically in New York and Eastern, and not at all in Northern and Western or Midland.

The Handbook of the Linguistic Geography of New England (Providence, 1939), by Hans Kurath and others, initiated a wider survey of a Linguistic Atlas of the U. S. A., which is on-going.

G. L. Brook (English Dialects, Oxford, 1963) has this to say about the diversity and historical connections of American English;

"One of the problems to be investigated by students of American dialects is the extent to which the English settlers in America brought with them dialectal variations that already existed in England. Professor W. N. Francis has pointed out that the Quakers who settled in Pennsylvania came from both Yorkshire and East Anglia, with the result that traces of both Yorkshire and Norfolk dialects are to be found in Philadelphia. Professor Francis further points out that it is possible to see a historical reason for the resemblances between British English and the New England dialect of American English in that the westward-moving pioneers severed their ties with England whereas the New Englanders did not. Hence New England kept up with the changing fashions of speech in England whereas the rest of the United States pioneer communities kept earlier pronunciations with little change. The dialects of East Anglia have often been suggested as the basis of American English, and this view accords well with what is known about the original homes of the earliest settlers, but it is doubtful whether we know enough about the dialects of British English in the seventeenth century to allow us to establish any but the most general links between British and American dialects." (p. 126)

In Canada, English is for the most part of the Northern variety of United States English, though there are areas where features of British English are maintained (especially in the Maritime provinces and Newfoundland).

In Australian English there is the same kind of phonological trend as in local Southeastern British English: vowels becoming more fronted, close, and often diphthongized. e.g. [səi:] sea, [yəu:] you. British [ei] becomes [ai], and British [ai] becomes nearly [ɔi].

PHONOLOGIES OF VARIOUS ENGLISH DIALECTS

The consonants of the various dialects of English differ mainly in minor points of phonetics and distribution. The following system needs to be supplemented only with a velar fricative /x/ for Scots dialects:

p		t	č	k	
b		d	ǯ	g	
f	θ	s	š		h
v	ð	z	ž		
m		n		ŋ	
		l			
		r			
w			y		

The vowel systems of the various dialects are quite diverse, both in the total number of phonemes and in their distribution and phonetic values, although it is not entirely certain that at least some of the diversity is not introduced by a great variability in analysis. At any rate, the differences among the Scots dialects themselves, for example, are considerably greater than the difference between Standard British and most varieties of American English.

A.C. Gimson (Phonetic Change and the R.P. Vowel System, in In Honour of Daniel Jones, David Abercombie et. al., eds., London, 1964) describes the vowel system of Standard British English as consisting of 5 long monophthongs / iː aː ɜː uː ɔː/ , 7 short monophthongs /ɪ e æ ə ʊ ʌ ɒ / , and 9 diphthongs / eɪ oʊ aɪ aʊ ɔɪ ɪə ɛə ɔə ʊə/ . In order to facilitate a greater degree of typological comparability we suggest a restatement of the above scheme as follows:

(Gimson's symbols)

iː	ɪ			uː	ʊ
eɪ	e	ɜː	ə	oʊ	ʌ
aː	æ			ɔ	ɒ

(Our symbols)

i		u
e	ə	o
a		ɔ

plus length.

None of the recent changes in this system appear to be thorough-going enough to alter it substantially: e.g. the diphthongization of certain vowels (both long and short), the monophthongization of certain diphthongs, and a few shifts involving either rounding or tongue height or length.

Herbert Pilch (Phonemtheorie, Bibliotheca Phonetica, Fasc. 1, New York, 1964) gives the following vowel system for Southeastern British English (without making mention of either diphthongs or length):

i	ə	u
e	ʌ	o
æ	a	ɔ

e.g. /pit/ pit, /pet/ pet, /pæt/ pat, /pət/ pert, /pʌt/ putt, /pat/ part, /put/ put, /pot/ port, and /pɔt/ pot.

The most common vowel system of American English (other than certain New England dialects) is as follows:

(Francis)

i		u
I		U
e	ə	o
ê		ɔ
æ		a

(Smith-Trager, after Bloomfield)

iy		uw
i		u
ey	ə	ow
e		o
æ		a

These systems are not to be construed as representing five absolute tongue heights, but rather only three relative ones since those vowels represented on 'levels' two and four are merely the lax and/or short counterparts to those on 'levels' one and three respectively. To the above inventory some investigators add a high central vowel /ɨ/ which presumably contrasts with all (or some) of the other vowels. Also, for some New England areas, the above scheme would have to be supplemented with at least one additional low vowel:

i u

e ə o

æ a ɔ

e.g., /bæt/ bat, /hat/ hat, /bɔt/ bought, /kôt/ coat, and /bot/ boat.

For many speakers of the northwestern United States (Carroll E. Reed, the Pronunciation of English in the Pacific Northwest, Language 37.559, 1961) the vowel system of Francis or Smith-Trager (given above) must be reduced by one member, since they lack the contrast between /a/ and /ɔ/. In some areas of western Canada the total vowel inventory is still smaller, since many speakers of this area do not contrast /æ/ with /a/, the two being allophones of the same phoneme. This system would then be:

i u
ɪ ʊ

e ə o
ɛ ɔ

a

In summary, Standard British has 14 vowels (short and long together), and North American dialects have from 10 to 12.

An excellent study of Scots dialects has been done by J. C. Catford (Vowel-Systems of Scots Dialects, Transactions of the Philological Society, 1957). His reason for choosing vowel systems (stressed, monophthongal vowels only, to be precise) is the fact that '... the greatest variation in Scots dialects is found in their systems of stressed vowels' (p. 117).

Catford's basic vowel system for all Scots dialects is:

i u
e ɪ o
ɛ ʌ
a

This system may be modified (thus increasing the total number of vowels) in any number of four ways:

(1) the addition of a high centralized rounded or unrounded vowel higher than /ɪ / (labeled Y),

(2) the addition of a fourth front vowel in the area of / e/ or /ɛ /; this vowel may contrast either in tongue height or length (variously in different dialects; labeled E),

(3) the addition of a low back vowel, rounded or unrounded in various dialects,

(4) the addition of a second low back vowel contrasting with the first in terms of rounding.

The basic eight-vowel system (without any of the above listed modifications) occurs only on the Isle of Bute, and in two other localities of which one is in

Ayrshire and the other in Lanarkshire.

Nine-vowel systems are very widespread. Those with an additional high centralized vowel (that is, the basic system of eight plus such a vowel) occur south of an isogloss line running across Scotland from about 20 miles south of Glasgow in the west, to just south of Edinburgh in the east. The nine-vowel systems north of this line consist of the basic system plus an additional low back vowel.

Berwickshire:

i	Y	u
e	ɪ	o
ɛ		^
	a	

Lanarkshire:

i		u
e	ɪ	o
ɛ		^
a		ɔ

The Berwickshire type of system occurs in Glasgow and surrounding areas.

Ten, eleven and twelve-vowel systems occur in three separated areas: in extreme southwestern Scotland; in a broad belt running north-south from the Firth of Forth to the English border (with nine-vowel systems both to the east and west); and in northern Scotland (i.e. north of the Firth of Forth right up to and including the Shetland Islands).

There are three types of ten-vowel systems:

N. Kirkcudbright:

e	Y	u
e	ɪ	o
a		ɔ

E. Fife;

i		u
e	ɪ	o
E		
ɛ		^
a		ɔ

A third type of ten-vowel system occurs in widely separated areas in southern Scotland:

i		u
e	ɪ	o
ɛ		ʌ
æ	a	ɔ .

There are three types of eleven-vowel systems occurring for the most part in southwestern and south-central Scotland:

i	Y	u
e	ɪ	o
E		
ɛ		ʌ
a		ɔ

i		u
e	ɪ	o
E		
ɛ		ʌ
æ	a	ɔ

i	Y	u
e	ɪ	o
ɛ		ʌ
æ	a	ɔ

Of the above three systems, the first also occurs in Kirkcudbright.

Finally, a twelve-vowel system occurs in N. E. Angus, the Shetland Islands and other northern areas beside several extreme southwestern areas:

i	Y	u
e	ɪ	o
E		
ɛ		ʌ
æ	a	

Note on orthography:: For low vowels we have used æ, a and ɔ when there is a threefold distinction; a and ɔ when there is a twofold distinction; and a when there is only one distinction. This orthographic regularization (which is also the case for the other vowel symbols) is not intended to show phonetic details, but is merely suggestive of typological regularities.

FRISIAN

Of all the Germanic languages, Frisian is most closely related to English.

Frisian is spoken by some 350,000 people living mainly in the province of Friesland in The Netherlands. All Frisian speakers are bilingual in Dutch, and Frisian is slowly receding in favor of the latter in spite of its use for official (governmental) purposes and for educational purposes (though the latter use is not very extensive). The declining use of Frisian is well illustrated by the fact that it is not spoken in the chief city of Friesland, namely Leeuwarden.

The above comments on the language situation in Friesland apply only to the West Frisian dialect.

Other Frisian dialects include
East Frisian
North Frisian
Island Frisian.

East Frisian is spoken in the northern part of the Oldenburg region by about 1,000 speakers. This dialect is in strong competition with Ostfriesisch Platt, a Low German dialect.

North Frisian is spoken in the Marsch region of the Schleswig coast by some 12,000 speakers. Subdialects include:
Gōsharde
Darrharde
Bökingharde
Wredingharde.

Island Frisian (excluding the West Frisian Islands which belong to the West Frisian dialect) is spoken by some 3,700 speakers in the North Frisian Islands (Sylt, Föhr and Amrum) and on the island of Heligoland. The decline of Island Frisian is indicated by the fact that nowadays investigators have difficulty in finding informants. The East Frisian Islands are inhabited by German speaking people.

About 30,000 North Frisian speakers have migrated to the United States, mainly New York and California.

The phonemes of Frisian (Cohen, Ebeling, Eringa, Fokkema and van Holk, Fonologie van het Nederlands en het Fries, The Hague, 1959) are:

p	t	k		i	ü	u
b	d	g		e	ö	o
f	s	x	h	ɛ		ɔ
	z				a	
m	n	ŋ		Plus length.		
	l					
	r					
w		y				

An additional vowel, /ə/, occurs only short. Phonetically, /g/ is [g] and [ɣ], /f/ is [f] and [v].

NORTH GERMANIC OR SCANDINAVIAN

The earliest known linguistic evidence for Scandinavian consists of runic inscriptions found in Norway, dating from around 200 A.D. The language of these inscriptions (till about 800 A.D.) is known as Proto Norse. This common Scandinavian language underwent considerable change between 500 and 700 A.D., for example umlauting, the loss of unstressed vowels (suffixes) and the loss of initial y and w as in /ɔ·r/ year and /ul/ wool respectively.

The period following this is known as the Viking Age or Old Norse, variously dated from 700 or 800 to 1050, 1250 or even 1350. During this time there developed enough dialectal divergence to permit a division of Old Norse into Eastern Old Norse (from which developed Danish and Swedish) and Western Old Norse (from which we have Norwegian, Icelandic and Faroese).

Middle Scandinavian (from Old to about 1525) saw a great number of innovations (notably the loss of the old case system) which lead directly to the Modern Period (1525 onwards).

During the Modern Period, Danish — for political reasons the most influential on all other forms of Scandinavian — has been in the forefront of linguistic innovation, followed by Swedish, then Norwegian, Faroese and finally Icelandic, the least innovating of all Scandinavian speech forms. In fact, Icelandic has undergone so little change that on the basis of lexicostatics — at least by one method of counting cognates — it may as well be regarded as identical with Old Norse from which it split off (geographically) in the 10th century.

Danish, Swedish and certain southeastern Norwegian dialects share a good number of the developments of the Old Norse period, including the monophthongization of certain diphthongs, e.g. stein > sten stone and øy > ø island. Norwegian and Swedish on the other hand share a number of features not found in Danish, largely due to their retention of a number of features which in Danish underwent subsequent innovation, e.g. the retention of unstressed 'full vowels' which in Danish and some southeastern Norwegian dialects were weakened to a 'slack e' (10th and 12th centuries, respectively) as in kastar > (Danish) kaster throws and visor/visur > (Danish) viser ballads. As is apparent from examples given so far, some southern Norwegian dialects share a remarkably large number of features with Danish, among them the shift from voiceless p t k after long vowels in syllable or word final position to voiced b d g, e.g. bok > bog book.

Gross morphological features shared by most Scandinavian speech forms include (after Priebsch) the presence of two genders, common and neuter, the suffixation of the definite article to the unaccompanied noun, and the use of the suffix -r to indicate all persons in the singular present indicative form of the verb.

In terms of well-established literary traditions there are six Scandinavian languages:

Icelandic (Iceland)
Faroese (Faroe Islands)
New Norse (Norway)
Dano-Norwegian (Norway)
Danish (Denmark)

Swedish (Sweden).

Icelandic was first written in an adapted form of the Latin alphabet around 1100. Faroese came to be established as a literary language through the efforts of V.U. Hammershaimb in 1850. New Norse (Nynorsk), called Landsmål before 1917, was introduced by the famous writer Ivar Aasen (1813-1896) in 1853 in an effort to overcome the extensive influence which Danish had come to have on Norwegian as a result of over 400 years of political union (1380-1814). This Danish influence on Norwegian had come to be perpetuated by the literary tradition known as Bokmål (called Riksmål before 1917, also called Dano-Norwegian). New Norse is based on those Norwegian dialects that are relatively free of foreign influence and most in accord with the linguistic traditions of the old classical period in Norwegian literature. Although there was strong opposition to the puristic-archaic New Norse at first, it received official recognition by legislation in the latter part of the 19th century. Bokmål is still used by the majority of urban Norwegians, but New Norse is gaining wider acceptance. Local school boards are free to decide on the use of either literary tradition as the medium of education. The disadvantages of two competing writing systems (actually much more than just writing systems, see below) in Norway have led to the introduction of a synthesis called Samnorsk, which, with the support of an official language committee since 1952, has the task of unifying New Norse and Dano-Norwegian. The language situation in Norway has further been affected by a series of three so-called 'language reforms' (1907, 1917 and 1938) which were not mere spelling reforms but which actually and substantially affected both pronunciation and morphological features (especially the 1938 reform). These reforms were intended to Norwegianize Bokmål and did not affect Nynorsk.

The situation as regards the amount of intelligibility between the various Scandinavian speech forms reflects their historical development in at least one very important sense: their various rates of retention of older features, or conversely, their rate of innovation of new features, Danish being the most innovating and Icelandic the least. From another point of view, namely the amount of contact, or conversely the amount of isolation, the situation with respect to intelligibility does not reflect the historical development of the Scandinavian languages very well, e.g. the large intelligibility gap between Norwegian and Icelandic which historically belong to West Norse, versus the high degree of mutual intelligibility between Norwegian and Danish, the latter historically a member of East Norse. E.I. Haugen (Encyclopaedia Britannica 20, 1964) sums up the situation as follows: "On the continent, Scandinavians are usually able to communicate with each other by speaking and writing their own languages. This is because of the common developments which the languages have undergone, including an extensive simplification of morphology and the adoption of loanwords from common sources, especially Low German. Danish is phonetically most deviant from the rest but has a large area of vocabulary in common with Dano-Norwegian; the cleavage is therefore greatest between Danish and Swedish."

In comparison with the Netherlandic-German speech area, continental Scandinavian certainly appears to be much more homogeneous.

Icelandic and Faroese, although not too different from one another, have become largely unintelligible to continental Scandinavian speakers due to their much slower rate of innovation, so that as far as intelligibility goes it appears reasonable to say that there are two Scandinavian languages, Insular and Continental.

The Scandinavian languages — traditionally, Icelandic, Faroese, Norwegian, Danish and Swedish — are today spoken by a total of some 18 1/2 million speakers, located for the most part in Scandinavian Europe. The only non-Scandinavian speakers of this area are Finns and Lapps.

(1) Icelandic is spoken in Iceland by 180,000 people (recent estimate by the Icelandic Embassy to the United States) and an additional 20,000 in North America.

Dialectally, Icelandic is the least internally diversified of all the Scandinavian languages. The main cleavage is between a Northern and Southern dialect, e.g. in the Northern dialect p t k are voiceless whereas they are half-voiced in the Southern dialect; Northern l m n before p t k are voiced (the stops being aspirated) where in the South l m n before p t k are voiceless (the stops being unaspirated); finally, Northern [kv] corresponds to Southern [xw]. There are also slight dialectal variations between farming and fishing communities.

(2) Faroese, with no more than 35,000 speakers, is the language of the inhabitants of the Danish-administered Faroe Islands located between Iceland and the Shetland Islands.

In contrast with Icelandic, Faroese is divided into practically as many dialects as there are inhabited islands, the number of which is seventeen.

(3) Norwegian is spoken by 3,600,000 people in Norway, and an additional 700,000 in the United States. Einar I. Haugen (Encyclopaedia Britannica 16, 1964) summarized the dialect situation in Norway as follows:

"Spoken Norwegian is divided into urban and rural dialects; the former has spread at the expense of the latter. Urban speech falls into standard and substandard social dialects. Standard urban speech is reasonably uniform throughout the country and serves as a model although some educated people prefer to speak a normalized New Norse or retain their rural dialects. The standard urban dialect is a compromise between traditional Norwegian speech habits and written Dano-Norwegian, as developed by the old official and professional class. Substandard urban dialects are closer to the surrounding rural dialects of each city, being historically the speech of rural-urban migrants. The rural dialects differ from parish to parish, but fall into broad regional types which reflect the paths of communication in medieval and early modern times: western (the fjord country from Romsdal to Setesdal), eastern (from Telemark to the Swedish border and north to the Dovre Mountains), Trönder (in the trading area of Trondheim), and northern (the three northernmost countries). New Norse has its strongholds in the western dialects, on which Aasen drew most heavily for his grammatical and lexical framework."

(4) Danish is spoken by 4,700,000 people in Denmark, by 400,000 in the United States, and by 10,000 in Germany. One of the most significant features that divides Danish into two main dialect groups is the so-called stød which is absent in the north and present in the south with Copenhagen roughly at the

boundary between the two areas. As a result there are speakers of both dialects in Copenhagen.

The stød feature, which is usually a glottal creak (and occurs with both vowels and consonants) corresponds to pitch-stress phenomena in Northern Danish, Swedish and Norwegian.

(5) Swedish is spoken by 7,500,000 people in Sweden, 1,000,000 in the United States and Canada, 400,000 in Finland (9% of the population), and a small number in Estonia. Swedish is thus the most widely spoken of all the Scandinavian languages, having also the largest number of speakers, approximately 9 million.

Swedish dialects can be divided into two main groups, each with a number of subgroups:

The Svea group including:

Uppland,
Hälsingland and other northern areas,
and parts of Östergötland.

The Göta group including:

Värmland
Västergötland,
parts of Smaland and other areas,
and the south Swedish provinces (which until long after the Reformation were Danish speaking).

SCANDINAVIAN PHONEMIC SYSTEMS

The phonemes of Icelandic as described by Einar Haugen (Language 34.

55-88, 1958) are:

p		t	k	
b		d	g	
f	θ	s		h
v				
m		n	ŋ	
		l		
		r		
			y	

i		u
e	ö	o
	a	

Plus length

In addition Haugen gives the following diphthongs most of which can also occur either short or long: / ii ei ai ui oi öi uu ou au / .

Old Icelandic (of about 1100) appears to have had the following phonemic system, attested on the basis of minimal pairs in an old piece of literature concerned with the modifications of the Latin alphabet needed to represent Icelandic (after Foster Blaisdell, 1963):

p		t	k	
b		d	g	
f	θ	s		h
m		n	ŋ	
		l		
		r		
w			y	

i	ü	u
e	ö	o
ɛ		ɔ
	a	

The vowel system of Danish is as follows:

i	ü	u
e	ö	o
ɛ		ɔ
	a	

All these vowels occur with a series generating component of length and another of stress (which may occur simultaneously). Another vowel, /ə/, constitutes a separate subsystem since it can be neither long nor stressed. The Swedish vowel system is practically identical with Danish, lacking only /ɔ/ (B. Malmberg, Jakobson-Festschrift).

The consonant system of Swedish and Norwegian (Pilch, 1964) is:

p	t	ṭ	k	
b	d	ḍ	g	
f	s	š		h
v				
m	n	ṇ		
	l	ḷ		
	r			

The retroflexed consonants apparently have developed historically from clusters of r plus t, d, n, l, as did š, which varies with rs.

EAST GERMANIC

The languages of this branch of Germanic are extinct. The dialect group including the Burgundians, Vandals, Gepidae and Rugii who were probably located in eastern Germany around 50 A.D. has left only traces and is consequently very poorly known. The Gothic group on the other hand is fairly well known, mainly from the Gothic Bible translation of 385 A D. by the Visigoth Bishop Wulfila.

The Goths are believed to have inhabited the area which is now modern Bulgaria. Speakers of Gothic may have migrated from Scandinavia prior to 50 A.D. The contention that the Goths came from Scandinavia is supported by the fact that there are a number of features which Gothic shares with Old Norse as against West Germanic. After 100 A.D. the Goths migrated south-eastward from the Vistula. Before 250 A.D. they had split into two groups on the plains north of the Black Sea, on opposite sides of the Dnieper. (Ostro-and Visi- refer to the east and west side of the river.)

A few words of 'Crimean Gothic' were collected in the 16th century from the remaining Gothic settlements. According to Bloomfield, some East Germanic settlements in the Crimea survived until the 18th century.

CELTIC BRANCH OF INDO-EUROPEAN

4.0. The name Celt (or Kelt) was used in its plural form (Keltoi) by Greek writers. The Greek reference was always to the Keltoi; the Galatoi or Galatians were Celts who had migrated into Anatolia from the Balkans.

Celts had spread not only to Anatolia (where they were known as the Galatians of the Bible), but also to Italy, and probably the British Isles by 400 B.C. Greek writers three centuries later were to refer to them as the Galatai (Galatoi) and Gauls. The etymological relation of Latin Galli(a) to French Gaule is obscure.

The original period of Celtic migrations was from East to West, like all other Indo-European migrations to Europe (circa 8th century B.C.). Special books have been devoted to the earliest relations between the Celts and Germanic peoples, for example, but such relations remain obscure. At a somewhat later period, the Celtic spread was halted near the Rhine by Germanic-speaking (Teutonic) tribes, and halted in Italy by Italic-speaking tribes. Subsequently the Teutonic tribes coming from the north, and the Roman armies and colonists from the south overran and settled the Celtic territory on the continent. Apart from those spoken on the British Isles, all Celtic languages—except the reintroduced Breton—were subsequently replaced by Latin and Germanic languages.

Still spoken languages of the Celtic branch of Indo-European, as well as historically attested Celtic languages, are placed in one or another of two groups, either Brythonic (Brittonic)—Gaulish now extinct, Welsh, Cornish now extinct, and Breton -- or else Goidelic (Manx on the verge of extinction, and Scottish and Irish Gaelic). The two group names have the following derivations: Brythonic from Welsh Brython, meaning Briton; and Goidelic from Old Irish Goidel, meaning Irishman.

This bifurcation of Celtic is made primarily on the basis of such sound-changes as /k^w/ of Indo-European becoming [p] in the Brythonic group, and remaining /k^w/ in the Goidelic group. The validity of a division based on this traditional example is sometimes questioned; Hamp, for example, speaks of the poor diagnostic value of the traditional gloss. But the Brittonic (Brythonic) bifurcation can still be stated as reflecting two dialects of Common Celtic. In other classifications Common Celtic is considered only as influenced by geographic distribution: Gaulish (not to mention 'Celtiberian') as representative of Continental Celtic, versus the rest lumped together--that is, designating the remaining languages as Insular Celtic. Breton is spoken on the continent today; but as an offshoot of ancient southwest British Celtic, the forebears of Breton speakers back-tracked from England.

(1) Gaulish became extinct about 500 A.D. (and doubtless had a very precarious existence for centuries before that), surviving today only in place names and personal names, divine names, graffiti, and in other inscriptions. Tacitus mentions that Gaulish and the Brythonic language of Britain were very similar; the differences in the words cited, however, lend support to Caesar's statement that dialect variations existed between the two, as well as within Gaulish itself.

(2) Welsh, spoken monolingually by no more than 100,000, and bilingually by 652,750 in 1957, is called Cymraeg by its speakers. Welsh is the descendant of the Brythonic dialect spoken by ancient Britons, a dialect of Common Celtic. Common Celtic was, morphologically, a fully inflected language, with word terminal markers for gender, number, case, mood and tense. In modern Welsh some of these have been lost as a result of phonological zeroing of word-final syllable. In intervocalic positions, unvoiced stops became voiced. By the 9th century A.D., Welsh manuscripts show that the language was at that time—as it had doubtlessly been earlier—distinct from other Celtic languages. Modern Welsh is spoken north of the Bristol channel in two major dialects. The Northern dialect is spoken in the counties of Anglesey, Carnarvonshire and Merionethshire; the Southern, in the counties of Cardigan, Carmarthen, Glamorgan and Pembrokeshire. There were 2,500 monolingual Welsh speakers in Patagonia (Chubut territory), South America, in 1891.

(3) Cornish, believed to be extinct since before the 1800's, was the native language of the people of Cornwall. It also developed from a dialect of the ancient Britons. Prior to this differentiation, there was a development of 'lenition' (voicing) of intervocalic unvoiced stops; and then the further change of medial and final [td] to [sz]. The influence of English was considerable ever since medieval times; today Cornish has been replaced completely by English.

(4) Breton is spoken by 900,000 native speakers in four main dialects: Tregovois (around the northern part of the peninsula, with the village of Trequier as its center);

Leonard (almost at the northern tip of the peninsula, with St. Pol-de-Leon village as the center);

Cornouaillais (on the southwestern portion of the peninsula, around the town of Quimper);

Vannetais (on the southern portion of the peninsula, around the town of Vannetais). Despite the passive resistance of the French government, Breton still flourishes; efforts are being made to combine the first three dialects into a literary language.

Although Breton is now spoken in continental Europe (on the Peninsula of Brittany, France), it is a member of the Brythonic group of Celtic languages. This language developed from the Celtic dialects spoken in Britain; it was brought to France by refugees from Anglo-Saxon' raids. There is no evidence that Breton was influenced by continental Celtic (Gaulish); later, it was certainly influenced by French, especially in its lexicon. In phonology—except for nasalization of vowels—and morpho-syntax (the use of the past particle in passive verbal constructions, and the use of the verb glossed to do as an auxiliary), Breton was closer to Cornish than to any of the other Celtic dialects. For this reason and because modern Breton is geographically opposite Cornwall, it seems certain rather than merely probable that it represents an early offshoot from Cornish. Both Breton and Cornish are dialects of Common Southwest British.

The following inventory, for the Leonard dialect of Breton spoken at St-Pol-de-Leon, gives a phonemicization by W.B.S. Smith (The Breton

Segmental Phonemes, SIL 4:3-4. 52-69, 1946) of data obtained by Sommerfelt in 1920; some Celtic specialists regard Sommerfelt's 1920 analysis to be sounder than the 1946 restatement which follows:

p			t		k		i	y	u
b			d		g		e	ae	o
f	F		s	š		h		a	
		v	z	ž		x	with nasality /V/		
m			n		ŋ		with length /V·/		
			l	ľ					
			r						

/f,F/ are both bilabial voiceless, but /F/ is 'shorter and less tense' than /f/.
/ľ,ŋ/ are rare, occurring only in word final position.
The problem—whether stress is predictable or not—is solved by analyzing semi-vowels as /ǐ, ě, j, aě, y̌, ǒ, w/ and stating the rule that "The strongest accent within a word falls on any vowel followed by the homorganic semivowel, otherwise on the next to the last vowel in the word."

(5) Irish Gaelic was spoken by over half a million people in 1961 (543,000); the total population of the Republic of Ireland (Éire) is 2,814,703. The 1961 census figure cited (543,000) may well represent wishful thinking and an enthusiastic government policy.

It must include everyone who has been exposed to a bit of school Irish, but the percent of those who ever uttered a sentence in the language would be quite low. There are only between 40,000 and 45,000 actual native speakers, and practically no adult monolinguals. Two years ago, a teenage Irish speaker from Aran Islands wandered through Dublin streets for half a day before encountering anyone who could speak any Irish at all. (See now, B.Ó. Cúir's Irish Dialects and Irish Speaking Districts.) English and Irish (Gaelic) are both official languages, and it is government policy to encourage Gaelic. A knowledge of the language is now a requirement for most civil service examinations.

Irish Gaelic is subdivided into four dialects:

Munster in the southwest (the basis for the standard Irish language);
Leinster in the southeast;
Connacht in the central zone;
Ulster in the north.

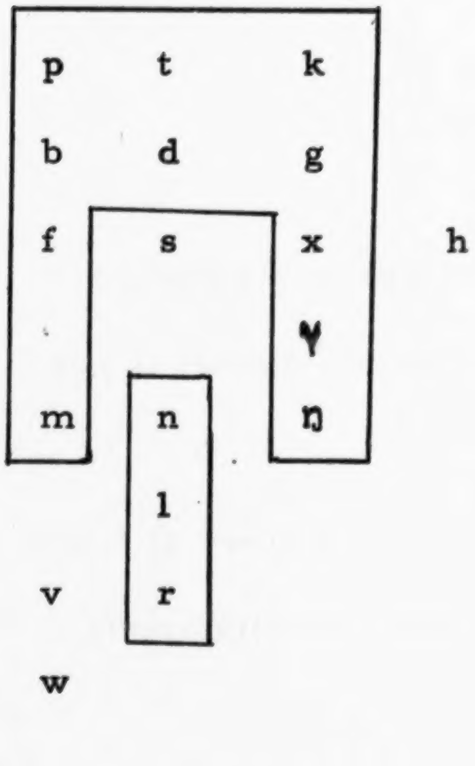
The 543,000 language census figure for 1961 given above represents speakers of the first three dialects; the number of speakers of the fourth dialect is not known.

Celtic linguists--especially those of Ireland--divide the history of Irish Gaelic into four periods (but see Kenneth Jackson, Language and History in Early Britain):

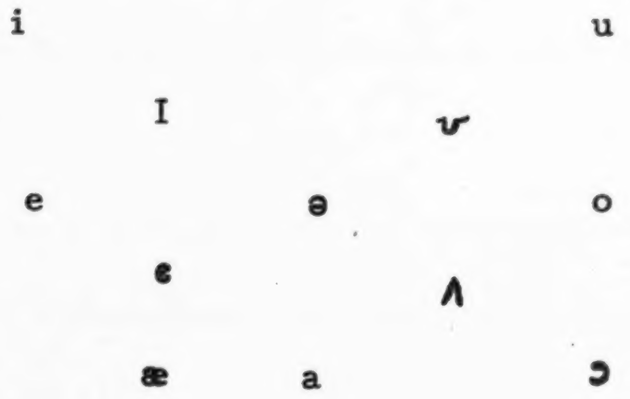
that of the Ogham inscriptions, probably between 300 A.D. and 600 A.D.;
Old Irish 600-900 A.D.;
Middle Irish 900-1200 A.D.--it was in this period that infixed object- and subject-pronominal markers were replaced by independent pronouns in the third person form of the verb.

Modern Irish, 1200 A. D. to the present day—it was in this period (about 1600 A. D.) that Scottish Gaelic emerged as a distinct language, developing from the Old Irish dialects of Irish speakers who had been migrating to Scotland since the latter half of the 5th century.

The particular variety of Aran Irish described by John P. Hughes (The Science of Language, N. Y., 1962, pp. 277-82) is spoken on the island of Inishmore, and is referred to as one of the Connaught (northwestern) types; the soundness of this analysis is questioned by some Celtic scholars.



/i u e o ə ɒ/ are described as tense; the others as lax



Stress is non-phonemic:

primary occurs with first syllable (all vowels except /ə/);

weak occurs with second syllable (only /ə/);

secondary occurs on third syllable (only /i u ə/).

Additive components that combine with consonants are labialization and palatalization, distributed as follows:

/n l r/ occur plain, labialized, and palatalized;

/s/ occurs plain and palatalized;

/h v w/ occur plain only;

all others occur labialized and palatalized, but not plain (except where neutralized, as in certain clusters).

In other Irish Gaelic dialects, /v w/ participate in the labialized-palatalized contrast.

(6) Scottish Gaelic diverged from Irish Gaelic at a late stage, as mentioned above; hence Scottish Gaelic shares many features with Irish Gaelic. There are, however, innovations which isolate the Scottish Gaelic dialects. The opposition between voiced and voiceless stops has been displaced; the reflexes in Scottish Gaelic are unaspirated and aspirated stops, respectively. Nasalization of vowels adjacent to a nasal continuant occurs in Irish Gaelic but not in Scottish Gaelic. Most of this exposition can be better stated in terms of a tense and non-tense contrast (rather than voiced-voiceless). The opposition ± tense is found in Irish also. As a mark of tenseness, the implementation of aspiration varies.

In Scotland, the total population was 5,178,490 in 1961. At that time 75,508 spoke Scottish Gaelic; 1,079 of these speakers were monolingual. The majority of the speakers are concentrated in the north central provinces—Ross and Cromarthy, and Inverness—and in the western off-shore islands—the Hebrides and the island of Skye. For children of Gaelic speakers, Gaelic is the language of the home; English the language of the school. The percentage of people speaking Gaelic is high among pre-school children, low among school children, and high among those above the age of sixty.

In the New World, Scottish immigrants formed a large proportion of the settlers of Nova Scotia and Cape Breton Island, where Scots Gaelic dialects are still spoken. Almost seven thousand Scottish speakers were reported for Nova Scotia in 1951.

Scottish and Irish Gaelic (and Manx) share some innovations which differentiate them from Welsh. Consonants have two series generating components, one of labialization and the other of palatalization, with the continuants /l, n, r/ showing an opposition of strong and weak forms. Vowels are nasalized adjacent to a nasal. These Gaelic languages have also borrowed lexical items from Brythonic languages and Latin. The Brythonic loanwords are recognizable because they show a distinctive Brythonic development from Celtic not reflected in the north (Scottish and Irish) languages. Loanwords from Brythonic to Goidelic are few in number and date from the immediately prehistoric period of the Irish language (1st through 5th century A. D.) before the differentiation of the Brythonic (Brittonic) languages, as is shown in reflexes or lack of reflexes from Common Celtic (by Calvert Watkins, personal communication).

(7) Manx, spoken on the Isle of Man, is now thought to be almost completely replaced by English. The estimated number of Manx speakers given by Meillet-Cohen (p. 54)—56,000 in 1954; 55,253 in 1951--must be for the total number of people on the Isle of Man (compare the 1961 population census figure of 48,150) rather than for Manx speakers. Kenneth Jackson estimated that only four or five speakers of Manx remained in 1955, and they were old persons; D. W. Greene (Professor of Irish, Dublin University) states in the 1964 Encyclopedia Britannica (5.150) that no native speakers now exist.

(8) Although some scholars include Pictish—formerly spoken in Scotland, and ousted by the better known Celts—among the Celtic languages, information on the origins and affinities of Pictish is slender; it is an extinct language, and remains unclassified.

THE SLAVIC BRANCH OF INDO-EUROPEAN

5. Except for a few border minorities, and North Americans with Slavic mother tongues, all the Slavic languages are located in countries with Slavic majorities and with official languages being Slavic: Poland, Czechoslovakia, Yugoslavia, Bulgaria, and republics in Soviet Russia. The total number of speakers exceeds 250 million.

Slavic has had a long history of linguistic contact with a number of other Indo-European languages. Summarizing the relationship of Slavic to the other languages, Roman Jakobson (Slavic Languages, 1955) says: "Common isoglosses testify a close and prolonged neighborhood with Germanic, Iranian, probably Thraco-Phrygian and foremost with Baltic, which is tied to Slavic by significant innovations both in vocabulary and in grammatical and phonemic, especially prosodic, features. Old loanwords from Iranian pertain mostly to spiritual, and those from Germanic to material culture. Contact with Altaic and Finno-Ugaric languages seems to be confined to the late stages of Protoslavic and has left but scanty vestiges in its vocabulary." G. Bonfante (Colliers Encyclopaedia, 1961) in demonstrating the connection of Slavic and Baltic, believes the affinity, while valid, to be overstated, and that very important Slavo-Iranian common features are overlooked, e.g. the change of /s/ to /š/ after /i/, /u/, /k/, and /r/, the locative plural *-su, the identical treatment of the Indo-European velars which become /s z ʐ/, and of the

IE aspirates, /bh/> /b/, /dh/ > /d/, the palatalization of the labiovelars before /e i/, and several important words such as bogŭ *God* (Old Persian baga-). Slavic also shows some connections with Tocharian and Armenian, as in the particle formed with *-lo-.

There are thirteen literary standards in Slavic, including both Macedonian (which has had a recognized literary form only since 1943) and Kashubian (whose literary form is not of great import). These are listed below:

Russian
Belorussian
Ukrainian
Polish
Kashubian
Lower Lusatian
Upper Lusatian
Czech
Slovak
Slovene
Serbocroatian
Macedonian
Bulgarian.

Between the nuclei formed by these literary norms there are scarcely any linguistic frontiers, since transitional dialects--both in terms of shared features and intelligibility--connect each area. Even across the non-Slavic belt of languages (i.e. German, Hungarian and Rumanian) separating those Slavic languages to the north from those to the south there are striking structural features, as well as a fair degree of intelligibility, as for example between Central Slovak to the north and Serbocroatian to the south.

In addition to this dialect chain which connects the Slavic languages, Horace G. Lunt (Encyclopaedia Britannica, 1964) calls attention to the fact that: "Almost any two intelligent Slavs can manage fairly quickly to establish elementary communication on a simple conversational level, for the languages have preserved a remarkable degree of uniformity in over-all grammatical pattern and in the vocabulary of every day life." (e.g. although the domains are restricted, Macedonian speakers were able, without difficulty, to understand Krushchev's broadcast speech before the U. N.)

In terms of more immediate intelligibility the Slavic literary nuclei can be grouped into three zones: East Slavic, West Slavic, and South Slavic. In terms of shared structural features this grouping also holds well, although Czech and Slovak have strong South Slavic affinities. Slovene, although strongly South Slavic in terms of shared features, stands apart from the other Slavic languages as not being a link in the intelligibility chain; in fact, various of the isolated dialects of Slovene are only marginally intelligible with each other (see below).

We now list the Slavic literary norms in terms of these broad zones of intelligibility, noting, however, that there are transitions between each zone:

East Slavic:

- Russian
- Byelorussian
- Ukrainian

West Slavic:

- Polish
- Kashubian
- Lower Lusatian
- Upper Lusatian
- Czech
- Slovak

South Slavic:

- Slovene
- Serbocroatian
- Macedonian
- Bulgarian.

To indicate the high degree of comparability between the various zones the cognate densities between several languages are indicated below from I. Fodor (The Validity of Glottochronology On the Basis of the Slavonic Languages; Studia Slavica VII (fasc. 4) 295-346, 1961):

	100 word list	Full list
Polish-Ukrainian	91	92
Polish-Bulgarian	91	91

Russian-Slovak	90	88
Russian-Lower Lusatian	83	83
Slovak-Lower Lusatian	89	89
Russian-Czech	90	89
Serbian-Croatian	98	98

Although it is customary to speak of three 'branches' in Slavic -- East, West, and South Slavic (based upon certain consonant combinations assumed to have been maintained until the loosening of Slavic unity) -- these do not perfectly represent the historical branching of Slavic. In the first place Czech and Slovak are regarded by many as being originally South Slavic, in the diachronic sense, even though the long period of contact with West Slavic has resulted in diffusion and levelling of features which now serve to identify Czech-Slovak with West Slavic synchronically. In the second place it is controversial as to whether there ever was a Proto West Slavic analogous to Proto East ('Old Russian') and Proto South Slavic. On the other hand, Proto East Slavic and its subsequent differentiation is fairly well understood, and Proto South Slavic (except for the Czech-Slovak problem and the peculiar position of Macedonian) and its subsequent differentiation has been sufficiently investigated so as to be no longer controversial.

The Slavic people began spreading from their 6th century homeland northwest of the Carpathian mountains until, in the 8th to 10th century, they were spread close to their present centers in a large contiguous area of Eastern Europe, the Balkans, and southwestern Russia, with Central Slovak at the very center. A later intrusion separated Slavic geographically into a northern and a southern area, but the isoglosses between Serbocroatian in the south and West Slavic, on the one hand, and between Bulgarian in the south and East Slavic, on the other, testify to an earlier contact.

In the northern area an important line of isoglosses separates Ukranian and Byelorussian from Polish, and Ukrainian and Polish from Slovak. In the South a similar isogloss line separates Bulgarian and Serbocroatian, with a fan of isoglosses, resembling the Rhenish Fan in Western Germany, spread through the Macedonian area.

The earliest extant Slavic documents are from late 10th or 11th century (a funerary inscription in 993). The majority are in Glagolitic script and are religious in content. The most important linguistically are the Gospel books Codex Zographensis and Codex Marianus. Two Chief manuscripts are also in Cyrillic, i.e. the Sava Gospel book (Savvina Kniga), and the Codex Suprasliensis.

The earliest known historically attested Slavic dialect is Old Church Slavonic. In 862 Prince Rostislav of the Great Moravian State appealed to Constantinople for Christian missionaries who would preach in the native language and train a native priesthood in order to counteract the growing influence of German missionaries. The Greeks sent the brothers Constantine (Cyril) and Methodius. Cyril and Methodius were natives of Salonika and spoke the Slavic dialect of that area; it was they who reshaped the Greek alphabet into the Cyrillic alphabet. The earliest extant manuscript dates, however, from a century later, and it is not possible to reconstruct the exact dialect used by the earlier disciples and their immediate successors.

When Old Church Slavonic became the liturgical language of the Bulgarian, Serbian, and Russian Churches, it came under the influence of the local dialects, and variants began to emerge: Bulgarian Church Slavonic, Russian Church Slavonic, and Serbian Church Slavonic, which continue in liturgical usage to the present day.

Three alphabets -- Glagolitic, Cyrillic, and Latin -- are used for the Slavic languages.

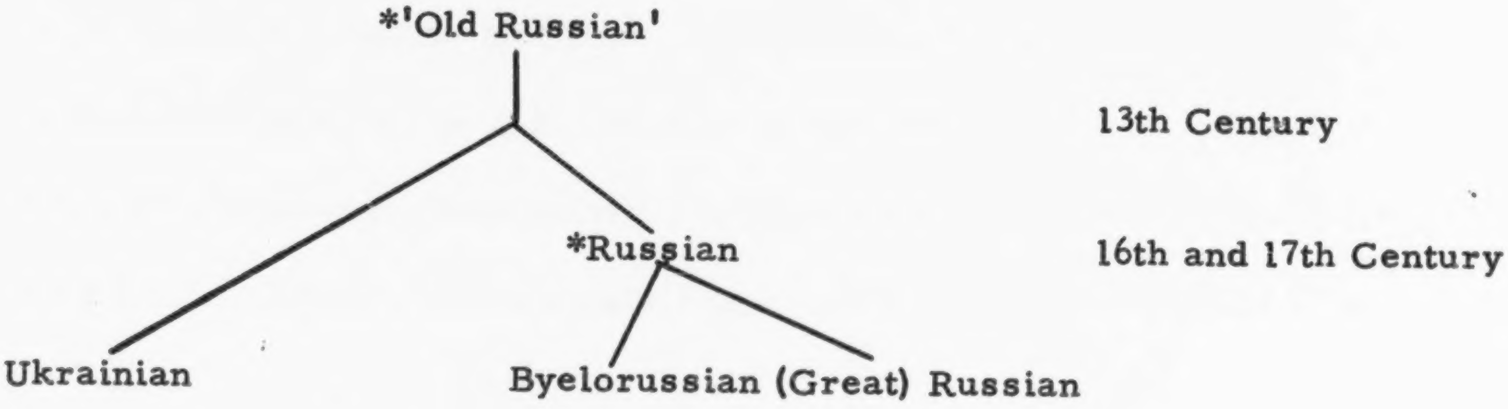
A form of the ancient Glagolitic alphabet, called Glagolitsa, was developed by the Croatians in the 14th century. Among Orthodox speakers, the Glagolitic was dropped; but it continued to be used among the Catholics of the Western Balkan Peninsula, where it was preserved in the Slavonic liturgy (in Dalmatian and Montenegrin communities).

Cyrillic was adopted by all Slavic peoples who accepted the Orthodox faith -- Bulgarian, Macedonian, Serbian, Russian, Byelorussian and Ukrainian. The differences among the various Cyrillic alphabets employed is not great: of the 45 symbols, there are 25 main letters with equivalent phonetic values in all the Cyrillic-using groups; five which are no longer used; six special symbols which are used only in Serbocroatian, three which are used only in Russian, and one which is used only in Bulgarian.

The Latin alphabet is used by those embracing Catholicism -- Poles, Czechs, Slovaks, Lusatians, Slovenes and Croatians. Latin characters were adapted to Slavic by some special conventions, e.g. by using a symbol otherwise not employed, such as c for [ts]; by diagraphs, cz for [č], and by the use of diacritics, as in the Polish nasalized vowels /ǫ, ę/. Some of these conventions have diffused to phonemic transcriptions for recording previously unwritten languages in Native America, Africa, and Oceania.

Before listing the languages according to their respective zones, we discuss briefly below some historical and comparative aspects of each zone.

The earliest records of East Slavic are preserved in 10th century chronicles. Although these are generally referred to as 'Old Russian', the texts represent an early stage of East Slavic when Ukrainian, Byelorussian, and Russian belonged to a homogeneous dialect area, with a chain of local varieties reaching from north to south. There are indications even in the earlier texts, however, of differentiation between a southern dialect group, and a northern dialect group, although the southern group did not become noticeably divergent until the Ukrainian split in the 13th century. Differences in the northern area became noticeable between an eastern and a western group until, in the 16th and 17th century, Byelorussian and Russian split. The following diagram represents this schematically:



The splits, however, were not carried out in discontinuous isolation, nor has there been geographic discontinuity since; the modern dialects of East Slavic still represent a geographically contiguous area. In fact, the features separating the three areas are maximally differentiated only in the central portions of these three major dialect areas (represented by the literary forms), with large 'zones of transition' -- local variants sharing features, to a greater or lesser degree, with more than one major dialect area -- connecting Byelorussian and Russian on the one hand, and Byelorussian and Ukrainian on the other. Such a transitional area does not exist, however, between Russian and Ukrainian.

Phonemically, the major differences between the East Slavic languages concern (1) the point of articulation at which there is phonemic palatalization -- labial, dental, palatal or velar; and (2) the number of stressed and unstressed vowel phonemes. If one were to ignore the complications which have been introduced by a number of loan words in each language, one would have the following situation for the three standard languages:

(1) For Russian all labials, dentals and /k/ are palatalized; for Belorussian all labials and all dentals but /t d/ and /r/, are palatalized; and for Ukrainian, dentals are palatalized and all labials and velars are not.

(2) Russian has three unstressed vowels and five stressed /i í é a á ó u ú/; Belorussian has three unstressed and five stressed vowels /i í é a á ó u ú/; and Ukrainian has six unstressed and six stressed vowels /i í e é ɪ ɪ́ a á o ó u ú/.

Lexically, Fodor indicates that, if one includes synonyms, the following scores of comparability are obtained from the 100-word and full Swadesh lists:

	100-list	full list
Russian and Belorussian	94	93
Belorussian and Ukrainian	92	91
Russian and Ukrainian	88	86

A study (in press) by B. L. Derwing and N. W. Schutz (using a 100-item list, translated into each language, and obtained by random selection (about every 50th word) from the alphabetical list in H. H. Josselson's The Russian Word Count; Detroit, 1953) indicates, however, that in a list which is not historically skewed Belorussian and Ukrainian appear to be much closer to each other than either is to Russian:

Russian to Belorussian	65
Belorussian to Ukrainian	82
Russian to Ukrainian	68

The results indicate that East Slavic represents a single language with a high degree of lexical and phonological sameness.

West Slavic may be divided into a Northern language (Lechitic, including Polish, Kashubian and the extinct Polabian), a West central language (Lusatian), and a Southern language (Czecho-Slovak), though this is not suggested by cognate density, including synonyms, given by Fodor:

	100-word list	Full list
Polish - Lower Lusatian	90	98
Polish - Upper Lusatian	89	89
Lower Lusatian - Upper Lusatian	98	98
Polish - Czech	93	92
Lower Lusatian - Czech	97	92
Upper Lusatian - Czech	94	92
Czech - Slovak	100	100

The earliest South Slavic split was between an Eastern and a Western group, with Bulgarian forming the Eastern group, and Serbo-Croatian and Slovenian forming the Western group.

The position of Macedonian is, however, undecided, as it shares on the one hand some features with Bulgarian, and on the other hand some features with Serbo-Croatian. The existence of some independent features and combination of features tend to indicate some independent developments. Macedonian is sometimes grouped in a Bulgarian-Macedonian zone as opposed to a Serbo-Croatian-Slovenian zone. Most investigators, such as Ivić, tend to support the view which ascribes independent position to Macedonian, between East and West Slavic.

The separation into an Eastern and a Western group is based by Lunt (Grammar of the Macedonian Literary Language, Skopje, 1955) on the development of the 'jers'. Ivić (Die Serbokroatischen Dialekte, 1958; On the Present State of the Study of Standard Macedonian, Word 9. 325-338, 1953) agrees with

Lunt on the Eastern-Western division, and adds a whole sheaf of isoglosses separating the two areas. The isogloss line runs in Western Bulgaria from a point just west of Vidin in the north, southward in a semi-circle between the Bulgarian border and the cities to the east of Sofija and K'ustendil, to the Macedonian border, where the isoglosses spread in a pattern resembling the Rhenish fan in Western Germany.

Linguistic criteria, if not the criterion of literature, allow us to see that all the South Slavic languages bear a strikingly close relationship. Not only is there a spectrum of gradations from one dialect to another in South Slavic, with accompanying mutual intelligibility, but the speakers of the standard languages are able to converse with one another. South Slavic, like East Slavic, is a single language.

Fodor gives the following percentages for the 100-word list and the longer list, including synonyms:

	100-word list	225-word list
Bulgarian and Macedonian	96	92
Bulgarian and Serbocroatian	92	90
Bulgarian and Slovenian	90	90
Macedonian and Serbocroatian	94	90
Macedonian and Slovenian	88	87
Serbocroatian and Slovenian	95	96

THE EAST SLAVIC ZONE

The three dialect groups of the East Slavic Zone are located almost entirely within the boundaries of the Soviet Union, primarily in the SFSR, the Ukrainian SSR and the Byelorussian SSR. The number of East Slavic speakers totals 180 million, 75 percent of whom speak Russian, 20 percent Ukrainian, and 5 percent Belorussian.

(1) Russian (Great Russian) is spoken by 136 million people primarily within the borders of the Soviet Union -- in the central area of European Russia, including Moscow and Leningrad; and in a great part of Siberia on both sides of the Siberian railroad, where speakers of Russian represent 83 percent of the population. There are 112 million in the SFSR, 8 million in the Ukrainian SSR, and 13 million in other SSR's as in the Kazakh SSR, where there are more speakers of Russian than of the indigenous language (Kazakh). In addition to being the native language of the largest number of people in the USSR, Russian is spoken as a second language by nearly all the speakers of other languages. There are about 3 million Russian speaking expatriates now living in North America, France, Germany, and China; half of these expatriates live in the United States.

Dialect variation is slight; the following gives the three main Russian dialect groups.

Central Group comprises the dialect spoken in a narrow belt running from northwest to southwest across the European part of Russia. The belt extends from Pskov to the Kalinin and Moscow area, and from a little north of Ryazan to Penza, and almost to Saratov on the Volga. The northwest portion bordering on Belorussian is sometimes classified as a separate subdivision.

Northern Group comprises the dialect covering the area to the northeast of the Central Group from Leningrad up to and including most of the Russian-speaking part of Siberia. The following are subgroups of the Northern group:

Olonets
Pomorsk
Leningrad-Novgorod (west)
Vologda-Kirov (east)
Vladimir-Volga

Southern Group comprises the dialect stretching southwest from the central region to the boundaries of Byelorussian and Ukrainian-speaking areas, and starting just south of Rzhev in the north. There is a transitional area between the southern dialect group and Byelorussian from north and south of Bryansk to near Rzhev. To the southeast the Southern group borders on Kazakh and Kalmyk territory around the lower Volga, and extends in a narrow strip to the Caucasus south of Stavropol' and Dzandzikan on the Terek, and includes Tula, Ryan-Tambov, and Orel.

The primary cleavage is between the Northern and Southern dialect groups, with the Central group sharing features of both. The Southern dialects have the phoneme /γ/ corresponding to the Northern /g/ (except in the Northern Pomorsk and Olonets dialects which have the adjective ending in -oγo as against other northern -ogo, -ovo, or ova); the Southern pretonic /a/ corresponds to the northern /o/, as in Southern /vadá/ versus the Northern /vodá/ *water*; the Southern uses of /-t/ *third person singular* and *plural* of verbs versus the northern /-t/ (except in the Olonets dialects). The Southern dialects have a /c/ versus /č/ contrast which does not exist in the northern dialects.

The literary standard based on the Moscow dialect (of the Central Group) shares some features of both the Southern and Northern dialects. Thus, it has the /c/ versus /č/ distinction and the 'akanie' pronunciation (/vadá/ rather than /vodá/) of the Southern dialects; but the plosive /g/, as in the Northern dialects, corresponds to the Southern /γ/.

Russian is the most widely-known Slavic language, with a classical literature read throughout the western world. Modern Russian is generally considered to date from the time of Peter the Great (eighteenth century) who introduced reforms to simplify the archaic church alphabet, and to encourage a simple, direct style of writing. Later in the same century Lomonosov investigated Russian styles, and advocated the adoption of the Moscow dialect as the standard. His influence was generally effective and was followed, at the turn of the 19th century, by a remarkable group of poets who helped establish the standard language by writing in it rather than in the archaic Church Slavonic. It was in this period that Alexander Pushkin -- often called both the Shakespeare and the Shelley of the Russian language -- had the effect of crystallizing the Russian literary language. The modern revised Russian Cyrillic alphabet was introduced in 1908, and modified slightly in 1917-18.

(2) Byelorussian (Belorussian, White Ruthenian, White Russian) is spoken by 38 million people in the Byelorussian SSR. The northern boundary is some 400 miles west of Moscow, and is roughly bounded by the Upper Volga in the northeast and the Desna River in the southeast, the Pripet River in the south, Narev River in the west, and stretches as far as the upper regions of the River Velikaya in the north (beyond the Dvina). About 33 of the 38 million Belorussians live in this area; another 4 million live in other SSR's of the Soviet Union; another million live outside Soviet Russia, especially in the U.S. and Canada.

Byelorussian has been divided into two or three main dialects. Karskij distinguishes two dialects according to the reflexes of *r'), and others make the same classification on the basis of the types of 'akanie.' A classification into three dialect groups has recently been proposed by N. T. Vojtovic (o dialektnoj osnove belorusskogo literaturnago jazyka, Voprosy Jazykoznanija, 1954, pp. 26-41). The more conventional groupings of the dialects is into the following:

The Northwestern Group, transitional to the Great Russian Southern Dialects;
The Southwestern Group, transitional into Northern Ukrainian.

Differences between these two dialects involve the following correspondences.

In the Northwestern dialect the pretonic /e/ and /o/ coalesce to /a/ when a high or mid vowel appears in the stressed syllable but to [ə~I] when /a/ occurs in the following stressed syllables (assimilative 'akanie'), while in the Southwestern dialect the pretonic /e/ and /o/ coalesce to an /a/ regardless of the vowel in the following syllable (dissimilative 'akanie').

In the Northwestern dialect /-c'/ *verbal 3rd person singular*, corresponds to the Southwestern zero.

In the Northwestern dialect /-có/, *verbal 2nd person plural*, corresponds to Southwestern /-cé/.

In the Northwestern dialect /-yosč/, *is*, corresponds to the Southwestern /-yesc/.

In the Northwestern dialect pretonic /ye, yo, ya/ coalesce to /ya/ if the tonic syllable contains a mid or high vowel, or to /yi/ if the tonic syllable contains an /a/ (assimilative 'akanie'), while in the Southwestern dialect pretonic /ye, yo, ya/ coalesce to /ya/ regardless of the vowel of the following syllable (strong 'yakanie').

Some varieties of the Northwest dialect have a palatalized /r'/.

In general, the Southwestern dialects contain many features in common with the northern Ukrainian dialects. It is the Southwestern dialect on which the Byelorussian standard language is based.

Byelorussian uses a Cyrillic alphabet beside a Latin alphabet which was revised to its present form in 1933. The Latin alphabet follows Polish conventions.

(3) Ukrainian (Little Russian) is spoken by 38 million people primarily in the Ukrainian SSR where it forms about 80 percent of the population. There are in addition large settlements of Ukrainian in other SSR's and in Europe (200,000) -- mostly Yugoslavia -- Canada (400,000), the United States (800,000), South America (Argentina, Brazil, Paraguay), as well as in Hungary, on the Lower Volga, and in Siberia.

Although Ukrainian is certainly East Slavic in classification, the long history of Ukrainian-Polish contact has left traces of many Polish loans. There are no clear linguistic boundaries between Ukrainian and Polish. The marginal dialects of the Ukrainian Carpathian dialects (called Rusnak) are strongly Slovak-influenced.

Ukrainian is divided into three dialect groups by some scholars, and into two by others.

The tripartite division distinguishes Northern from Southern from Carpathian. In the Northern Ukrainian dialects there is no /u/ phoneme; [I] and [ə] may or may not coalesce; there is final voicing of stops, but no palatalized liquid /r'/. In the Carpathian group there is an /I/ versus /i/ contrast; final unvoicing of stops; and Carpathian /t', k'/ correspond to other Ukrainian /k', g'/, respectively. The Rusnak dialect of the Carpathian group has some distinct characteristics, such as predictable stress (on the penult). In the Southern Ukrainian dialects there is no /u/ phoneme; no /I/ versus /i/ contrast; there is final voicing of stops, and also palatalized liquid, /r'/.

The bipartite division of Ukrainian into an Eastern dialect group and a Western group is derived from another set of differences:

Eastern, in the Great or Dniepr Ukraine;

Western, in Galicia, Podolia, Polesia, and Volhynia (formerly in Poland); and in a strip in northern Bukovina, formerly in Rumanian territory; and in the Transcarpathian Ukraine (formerly in Czechoslovakia, known as Subcarpathian Russia).

Eastern Ukrainian, the basis of one literary language -- while another literary language is based on the Southern dialect around Poltava and Kiev -- has more features in common with Great Russian than the Western dialect, as the latter is heavily influenced by Polish, and its dialects form a transition into Polish (e.g. the Eastern /s z/ become /ś ź/ nearer to Polish). There are many vocabulary borrowings and even common features in morphology (e.g. in some dialects of Ukrainian the past tense of the verb takes personal endings as in Polish).

Ukrainian literature began with the publication of Kotlyarevsky's 'Aeneid' in 1798, written in a colloquial style. The Standard language was established through the works of Kvitka-Osnovyanenko, Methynsky, Hrebinka, Kulesh, and Shevchenko. There was a period when the printing and writing of Ukrainian was banned under Czarist Russia, since Russians then resisted acknowledging the status of Ukrainian as a separate language; instead it was then regarded as the 'little Russian dialect'. Since 1917 Ukrainian has been the official language of the Ukrainian SSR.

THE WEST SLAVIC ZONE

There are four West Slavic dialect groups, if one counts Polish (1) as separate from Kashubian (2); Lustian (3) as a single dialect group (despite the existence of two literary norms); and Czech-Slovak (4) as a single dialect group. The nearly 50 million speakers of West Slavic are located primarily in Poland, Czechoslovakia, and East Germany.

(1) Polish is spoken by some 32 million people living for the most part in Poland; in addition there are 100,000 Polish speakers in the Těšin district of Czechoslovakia, well over 500,000 in the USSR, nearly 100,000 in Germany, 10,000 in Rumania, 130,000 in Canada, and almost 3 million in the United States.

The earliest documents relating to Polish -- proper names in Latin texts -- date from the 12th century; the earliest with connected Polish texts date from the 14th century.

Polish shows much foreign influence, most particularly in its vocabulary which has been enriched by Czech, Latin, German, Italian, French and English loans and flooded with Ukrainian and Byelorussian loans.

Although five dialect groups are customarily enumerated for Polish, one of these (Kashubian) is as diverse from Polish as is, say, Lusatian, hence, Kashubian is listed separately below (2), coordinately with Lusatian (3). Wielkopolska-Kujawy Group, centering around Poznań in Wielkopolska and around Inowroław and Włocławek in Kujawy;

Małopolska Group (Little Polish), centering in Kraków (Cracow);
Silesian Group, with Katowice as the main center;
Mazovian Group, centering in Warsaw.

(2) Kashubian (Cassubian) is spoken by 200,000 people on the left bank of the Lower Vistula in north central Poland (on the seaboard west of Gdansk (Danzig) running inland in a narrow strip toward the southwest from Gdynia). Kashubian was alternatively under Polish and German domination, and has had strong influence from both. It is heavily Germanicized. At an earlier period Kashubian shared closer relations with the northernmost Polish dialects than with Great and Little Polish; but Kashubian came under heavy German influence, while the northern dialects came under heavy Polish influence. At the present time Kashubian is being assimilated into Polish. Only a few folk publications, radio programs, and plays are written and given in Kashubian today. Slovincian is an archaic variant of Kashubian reported to be still spoken to the northwest of the compact Kashubian area. Around Fódź (except to the west) there are transitional dialects between Kashubian proper, Slovincian and Polish.

(3) Lusatian (Wendish, Sorbian) is spoken by about 150,000 people located in East Germany. Their area borders northwest Czechslovakia, and southwest Poland. The area of their land extends east of a line from Berlin to Dresden; it is centered for the most part on the upper reaches of the River Spree (Sprjewa). The eastern border runs from Görlitz (Zhorjelc) northward along the east bank of the River Neisse (Ńiža) and then along the Oder

as far as Fürstenberg (Přiběh), and finally west along the canals to the Spree River. This river forms the northwest boundary up to a point a few miles north of Lübben (Lubin). The border from this point lies west of Lübben, Lalau (Katawa), Ruhland (Rólany), Königsbrück (Kinsbórk), then runs southeast to Bischofswerda (Biskopicy) to the Czech border ten miles northwest of Šhuknov. It follows the Czech border to Zittau (Žitawa), and then northwest to Gorlitz. Upper Lusatia begins south of Spremberg. The urban areas include a majority of German speakers (80 percent), but in the rural areas Lusatians predominate (75 percent). The area has vacillated between Polish and German domination since it was conquered in 938.

Modern Lusatian has two literary forms corresponding to the independent political divisions of Upper and Lower Lusatia:
Upper Lusatian (High Sorbian, Upper Wendish, Hornju Lužica), centering in Budyšin;
Lower Lusatian (Low Sorbian, Lower Wendish, Dolna Lužica), centering in Kottbus (Chośebuz).

There is a third dialect group, beside the two literary forms:
Eastern Sorbian, around Muskau (Mužakow).

The Lusatian dialects are very similar, differing primarily in phonology. The speakers understand each other on very short contact. Lower Lusatian is in some respects rather like Polish; in other respects Lower Lusatian is least like any other Slavic tongue. Upper Lusatian resembles Czech, and bears even closer resemblance to Slovak and Old Czech.

Both dialects used Latin script until the 17th century, when the Gothic script was adopted. Catholic writers in Upper Lusatia have always retained the Latin alphabet. The oldest Lusatian document is from the late 15th century.

All Lusatians are bilingual with German as a second language. German has greatly influenced the Lusatian language, but Lusatian is still being taught in the primary and secondary schools. The present day Lusatian groups are the remnants of a large group. During the Middle Ages, a more extensive Lusatian area was flanked by the now extinct Polabian language to the north, Polish to the east, and Czech to the south, forming an intermediate zone between these three languages.

(4) Czech is spoken by 10 million people; it is about the only language spoken in the western part of Czechoslovakia (in Bohemia, Moravia, and Silesia). Small Czech communities in adjacent Kladsko and Upper Silesia are also found in present day Poland.

The relationship between Czech and Slovak is controversial only because of politico-cultural considerations. Linguistically it is clear that Czech and Slovak are so extremely close that they constitute a single language without question. Slovaks have adopted a literary form which differs from Czech, though not greatly so. In terms of intelligibility, free conversation is possible, though some slight interference is caused by the Czech vowel mutation.

Czech is divided into six dialect groups:

The Central Group, the basis of the Czech literary language, is centered in Prague (central Bohemia).

The Northeastern Group is located northeast of the Central dialect area (northeast Bohemia).

The Southwestern Group is located southwest of the Central dialect area (southwestern Bohemia); the Northwestern and Southeastern groups are geographically separated by the Central dialects.

The Czecho-Morvian Transition Group.

The Hanák Group (Moravia) is centered in Brno and Prostějov.

The Lach Group (Yalach, Northeastern) is located in Silesia and forms a transition to Polish on the one hand and to Eastern Slovak on the other.

(5) Slovak is spoken by over 4 million speakers, of whom a half million reside in the United States. The bulk of the Slovak population is located high in the Tatra Mountains, and in the surrounding foothills north of Hungary, in the eastern portion of Czechoslovakia.

Slovak might well serve as a Common Slavic lingua franca, since it shares many features with East Slavic, South Slavic, and West Slavic.

Three main Slovak dialect groups are distinguished; Eastern, Central, and Western. The Central dialect, on which the Slovak literary language is based is the most distinctive; it possesses certain features in common with South Slavic. Eastern and Western Slovak are closer to each other than to the Central dialect. Eastern Slovak shows certain affinities with Polish (loss of long vowels, and some common consonantal features). All dialects of Czech and Slovak are mutually intelligible--but Eastern Slovak and Bohemian Czech are perhaps somewhat less immediately intelligible to the other Czech-Slovak dialects.

SOUTH SLAVIC

There are four South Slavic dialect groups which number altogether about 25 million speakers. More than half of the speakers are Serbocroations; over a quarter are Bulgarians; in addition, there are two million Slovenians, and a million Macedonians.

(1) Bulgarian is spoken by about 7.8 million people located principally in Bulgaria. Of all of Slavic, Bulgarian stands apart with Macedonian in having lost its case distinctions (although vestiges of the cases exist in the oral tradition, and some dialects preserve these distinctions). The dialects of Bulgarian do not lend themselves readily to classification. There are three possible ways of dividing the dialects, according to R.G.A. De Bray (Guide to the Slavonic Languages, London, 1951).

A tripartite division based upon the phonological developments of Common Slavic /ě/ gives us the following division:

A Western Group, spoken west of an oblique line running from N.N.E. to S.S.W. from a little west of Nikopol on the Danube to the Greek frontier, with an eastward bulge around Cepino (Common Slavic */ě/>/é/);

A Northeastern Group, spoken east of the western group and north of a line running east from a little south of Pazardžik to the Black Sea south of Burgas (Common Slavic */ě/>/á/ or /é/, depending upon the phonological environment);

A Southeastern Group is spoken in the rest of Bulgaria south of the Pazardžik and east of the Western Group (Common Slavic */ě/> varieties of /á/).

One bipartite division is based on the developments of Common Slavic */tj; dj/; this gives us:

The Western Group spoken west of Sofia and into the northwest (*/tj/> /č/, and */dj/> /ǯ/;

The remaining dialects are spoken throughout Bulgaria (*/tj/>/št/ and */dj/>/žd/.

Another bipartite division is based upon the developments of Common Slavic /ǫ/:

The Central Group, which is spoken in a large area of Central Bulgaria east, southeast, south, and southwest of Sofia; this area extends into Greek Thrace (*/ǫ/>/á/);

The remaining dialects spoken throughout Bulgaria (*/ǫ/>/ə́/).

The Standard division for South Slavic however, is as follows:

Eastern Group

Western Group

Central Group.

The Modern Literary Language is based principally on the Eastern Group. Earlier writers had also used eastern dialects. Bulgarian is written in a Cyrillic alphabet.

(2) Macedonian is spoken by about a million people in the Macedonian People's Republic (autonomous unit in Yugoslavia), Bulgaria and Greece.

Macedonian, along with Bulgarian, has lost the Slavic case system.

There are six Macedonian dialect groups:

The Western group of dialects including:

the central dialect spoken in the Veles, Prilep, Kičevo, and Bitola regions in southeastern Yugoslavia;

the dialect spoken in the Debar-Galčnik region of southeastern Yugoslavia.

The Southwestern dialects are spoken mainly in Northern Greece in the Kostur and Lerin regions.

The Southeastern dialects are spoken in the region of Gevgelija, Strumica and Lake Dojran in Southern Yugoslavia.

The Southern dialect is spoken in the Kukus and Voden region of Northern Greece.

The Eastern dialect is spoken in the Štip and Pirin region of Southwestern Bulgaria.

The Northern dialect is spoken in the Kumanovo-Kratovo region of Southeastern Yugoslavia.

Macedonian has become a literary language only in recent years. During the struggle for liberation by the Yugoslav peoples in World War II, Macedonian was recognized as a national language (in 1943). The literary language is based on the central dialect (of the Western group). The poets Kosta Racin, Kole Nedelkovski, and Venko Markovski laid the basis for this dialect as the literary language. Macedonian uses a Cyrillic script.

(3) Serbocroatian is spoken by over 12 million people in Yugoslavia, in

the four Republics of Serbia, Montenegro, Bosnia-Hercegovina, and Croatia; and it serves as an important second language in Slovenia and Macedonia.

The dialect groups of Serbocroatian are conventionally referred to using as a label the form for the pronoun what (što, kaj, ča). To these što, kaj, and ča dialects, Ivić (Die Serbo-kroatischen Dialekte, 1958) adds a fourth group, Torlakian.

Štokavian (the što dialects) are defined by a combination of criteria, chiefly the development of the Common Slavic */ě/ (reflexes either of /je/ or /e/), and the degree of innovation in the accent system. These dialects are located in the central and eastern parts of Yugoslavia, covering the greater part of the country. There are eight sub-groups included by Ivić:

East Hercegovina
Šumadija-Vojvodina
Late Ikavian
Zeta-Lovćen
Kosova-Resava
Istrian Ikavian
Rumanian dialects with /i/
Slavonian

Čakavian (the ča dialect) is spoken in the Northern Dalmatian area of Western Yugoslavia, and on some Adriatic Islands.

Kajkavian (the kaj dialect) is spoken in Northwestern Croatia in Northern Yugoslavia, and is transitional to Slovenian.

Torlakian, spoken in the extreme southeastern Yugoslavia on the border of Macedonia, Bulgaria and Albania.

Ivić indicates that studies are being completed in 1,000 Serbocroatian dialects, the results of which are to be published in 1967.

The Serbs and Croats speak a single language, but because of nationalistic antagonisms minor regional characteristics are exaggerated in order to prove that they are separate languages. The most striking difference is not linguistic but orthographic: the Catholic Croats use a Latin script, and the Serbians use a Cyrillic alphabet. Although the Croats speak a western dialect, not all Serbs speak an Eastern dialect. That is to say, the dialect boundaries do not completely coincide with political and national feelings. Certain vocabulary items have been pointed out as being either Serbian or Croatian, but almost all such words have cognates in the other and are understood throughout the area.

The modern literary language is based upon the Štokavian dialect which was agreed upon in 1850 by a meeting of all the leading Yugoslav scholars. One of these scholars was Vuk Stefanović Karadžić, who worked out a Cyrillic alphabet and a Standard Serbian based on Štokavian (a'je-dialect' version); another scholar was Ljudevit Gaj, who worked out a Latin alphabet and standard based also on Štokavian (an 'e-dialect'). As a result Serbs and Croats use the same literary standard either with Latin or Cyrillic alphabet, or in the 'e- or je-variants. Strictly speaking it is the Orthodox Serbs and Bosnians who use the Cyrillic alphabet and the 'j-' version, and the Catholic

Croatians who use the Latin script, and the / e/ version. There is also a third group, the Dalmatians, who, although also Catholic, still to some extent use the ancient Glagolitic script for their Slavonic Liturgy (the only Catholic group to maintain the ancient liturgy).

(4) Slovene (Slovenian) is spoken by about 1.7 million people in Yugoslavia in the extreme northwest, in Carniola, in the southern parts of Styria and Carinthia, and over the Italian border in the province of Udine and the Vale of Resia. There are about 180,000 Slovenes in the United States.

The earliest record of Slovene is the Freising manuscript dating from around 1000 A.D., which contains a confessional form, a short homily and a confessional prayer. During the Reformation, Primož Trubar and a group of Protestants wrote in Slovenian, but their work was destroyed early in the 17th century during the counter-Reformation. Their orthography served as a model for the few Catholics who wrote the language in the next 200 years.

The Slovene dialects which are extremely diverse, developed during a thousand year period characterized by a lack of political and cultural unity. The villages in the Alpine regions of the north and west are particularly isolated and diverse. Included are two fairly large dialect groups:

The Lower Carniola.

The Upper Carniola.

The literary language is a compromise between these two, based primarily on the local dialect of Dolenjsko, with vowels borrowed from the local dialect of Gorenjsko. Slovenes tend to pronounce the literary language, however, with the vowels of their own local dialect when speaking formally; and in informal conversation they often use local grammatical forms. Horace G. Lunt (Encyclopaeda Britannica, 1964) reports that although there are gradual transitional dialects from Serbo-croatian to Slovenian, the two literary languages are more distant. Most Slovenes have some Serbo-Croatian in school and have at least a passive knowledge of it; but a Serb or Croat has difficulty reading and even more difficulty and little understanding when Slovene is spoken.

SLAVIC PHONOLOGIES

The following general sources were used for a number of Slavic languages: E. Stankiewicz, Towards a Phonemic Typology of the Slavic languages in American Contributions to the Fourth International Congress of Slavicists, (Moscow, 1958); H. Kučera, Inquiry into Co-existent Phonemic Systems in American Contributions to the Fourth International Congress of Slavicists; C. E. Bidwell, Historical Phonologies of Slavic languages in Tabular Form (The Hague, 1963); and H. C. Lunt, manuscript on Slavic phonemes. The sources used for particular languages will be cited below.

Every Slavic language includes consonants which involve the palatal point of articulation in one of three contrastive patterns:

(1) palatalized versus non-palatalized consonants in a sharp versus plain opposition (Russian, Belorussian, Ukrainian, and in eastern Bulgarian as one possible analysis of eastern Bulgarian);

(2) plain palatal (rather than palatalized in contrast with non-palatalized consonants)--that is, at a given point of articulation, a given consonant is unmatched by a palatalized opposition at the same point of articulation (Czech, Slovak, some western Bulgarian as the only possible phonemicization, Slovene, and Serbo-Croatian);

(3) both (1) and (2) above--that is, the palatalized versus non-palatalized contrast for some consonants, and the palatal articulation unmatched by non-

palatalized for other consonants (Polish and Macedonian).

Citation of languages in parentheses for each of these three patterns of the universal Slavic phonological involvement with the palatal point of articulation serves to identify the occurrence of a particular pattern rather than all its occurrences.

RUSSIAN

The following sources were consulted for Russian (in addition to the general sources listed above): R. I. Avanesov, Fonetika Sovremennogo Russkogo Literaturnogo Jazyka (Moscow, 1956); William S. Cornyn, On the Classification of Russian Verbs, Lg 24.64-75 (1948); H. I. Aronson, Morphophonemic Patterns in the Bulgarian Inflection, Ph.D. thesis, Indiana University (1961); G. L. Trager, The Phonemes of Russian, Lg 10.334-44 (1934); D. Ward, Is there a phoneme γ in Russian?, Le Maître Phonetique 112.29-31 (1959).

The following inventory of Modern Standard Russian is one that all sources would recognize as showing phonemic contrasts.

p	p'	t	t'	c	č	k
b	b'	d	d'			g
f	f'	s	s'		š	x
v	v'	z	z'		ž	
m	n'	n	n'			
		l	l'			
		r	r'			
					y	

i	u	í	ú
a		é	ó
		á	

Voiced stops do not occur finally.

Bidwell, Lunt, Aronson, and Kučera (but not Trager, Cornyn or Ward) include /k'/ as an additional palatalized phoneme (based on a few contrasts only). Bidwell, Lunt and Stankiewicz include /g'/ as a phoneme (though Stankiewicz considers it 'marginal'), while Trager and others analyze /g'/, like /k' x'/, as an allophone of the corresponding velar. Only Bidwell considers /x'/ contrastive. Trager, Avanesov, Aronson, and Ward include /γ/ as an additional fricative phoneme. The phoneme /γ / is recognized by all as contrastive in certain dialects.

Bidwell recognizes the possible contrast of [ʒ ǯ] , which is generally considered non-phonemic. Aronson includes /ʒ / as a marginal phoneme, and Cornyn includes /ǯ/.

Although unstressed [e] and [o] are not considered phonemes, they occur in contrastive positions in words identified as loans (Kučera). Cornyn includes /e/ as a phoneme, since it does occur in certain suffixes (even in native words).

BELORUSSIAN

For the Minsk Standard dialect Harold L. Klagstad gives the following inventory:

p	p'	t	c	c'	č	k	i	u
b	b'	d		ʒ'		(g)	e	a
(f)			s	s'	š	x		
v	v'		z	z'	ž	γ	í	ú
m	m'	n		n'			é	ó
		r			l		á	
					y			

/f/ and /g/ occur only in loanwords.

Voiced stops do not occur finally.

Bidwell includes the palatal stops /ć ȥ́/ rather than the palatalized /c' ȥ'/.

Stankiewicz and T. P. Lomtev (Byelorussian Language, 1956) include /r'/ in the inventory for Belorussian, but other sources exclude it or consider it marginal or dubious; it does not occur in the Minsk Standard dialect.

Other sources also include /l'/ and /ȥ̌/.

Stankiewicz and Klagstad exclude /ȥ/ from the inventory, but others include it; it occurs only in loanwords.

Some sources regard a contrast between /k x γ/ and /k' x' γ'/ as having been introduced by loans; in native vocabulary /k' x' γ'/ occur before front vowels /i e/, and /k x γ/ elsewhere.

The phonemic status of unstressed /e/ is uncertain, and many sources indicate a three-vowel system for unstressed vowels.

UKRAINIAN

The following phonemic inventory and comments for Ukrainian are from H. L. Klagstad, on Kiev Standard, and O. Panjko (Gramatika Ukrajiuskoji Movy, 1949), in addition to the general sources indicated in the introductory comments above.

p	t	t'	c	c'	č	k	
b	d	d'		ẓ'	ẓ̌	(g)	
(f)			s	s'	š	x	h
			z	z'	ž		
m			n	n'			
			r	r'	l		
					y		

i	ɨ	u
e	a	o

plus stress

/f/ and /g/ occur only in loanwords.

In addition to syllabic allophones, /u/ has non-syllabic allophones [w] and [v].

We write barred /ɨ/ where Klagstad writes small cap /I/; the phonemicization is the same whichever letter is used.

In some dialects unstressed /ɨ/ and /e/ do not contrast. So also, in some varieties, /o/ and /u/ are sometimes not in contrast in unstressed position. Thus the minimal unstressed vowel system is:

i	u
e	a

/t' d' s' z' c' ẓ'/ vary phonetically from much like Russian palatalized dentals (in the eastern dialects and the standard) to much like Polish palatal stops /t́ d́ ś ź ć ẓ́/ (in the western dialects).

Voiced stops occur word finally in Ukrainian, unlike Standard Russian and Belorussian.

Some sources add /ẓ/ to the inventory; some do not include /r'/; some have

/ɣ / rather than /h/.

Bidwell in addition indicates /k' x' ɣ'/ as phonemic and Klagstad suggests their 'marginal' phonemic status.

POLISH

The following sources were consulted for Polish in addition to the general sources indicated above: George L. Trager, La Systématique des Phonèmes du Polonais, Acta Linguistica 1.179-188 (1939); Zbigniew Folejewski, The Problem of Polish Phonemes, Scando-Slavica 2.87-92 (1956); Philip Scherer, Juncture in Polish, Lg 22. 353-8 (1946); J. Krotovskaja and B. Gol'dberg, Praktičeskij Učebnik Pol'skogo Jazyka (Moscow, 1959); Fransiszek Lyra, English and Polish in Contrast, Ph.D. thesis, Indiana University (1962).

Most of the sources agree in showing the following phonemic distinctions:

p	p'	t	c	ć	č	k	k'				
b	b'	d	ʒ	ʒ́	ǯ	g	g'	i	u	ę	ǫ
f	f'		s	ś	š	x		e	o		
v	v'		z	ź	ž			a			
m	m'		n	ń							
			l	ĺ							
			r		y						

Trager, Bidwell and Scherer phonemicize the palatalized consonants as clusters of consonants plus /y/, rather than as unit phonemes as shown above.

/e/ is said to be a 'potential' phoneme, which is omitted in colloquial standard (Lyra and Lunt).

An additional vowel phoneme, /ɨ/, is included in the inventory by Lyra, Bidwell, Scherer, Krotovskaja and Gol'dberg.

An additional palatalized phoneme, /x'/ is given by Folejewski.

Two less palatalized phonemes are given by Krotovskaja and Gol'dberg, who do not include /k'/ and /g'/ in the inventory.

For the Mazurzenie and Siakanie dialects, the phonemic inventories given by Lyra include the same number of phonemes, since the addition of /w/ to the inventory is offset by the removal of /l'/ from the inventory.

KASHUBIAN

Harold L. Klagstad gives the following inventory of phonemes for northern Kashubian:

p	p'	t	c	ć	č	k	i	ə	u		
b	b'	d	ʒ	ʒ́	ǯ	g	e		o	ę	ą
f	f'		s		š	x		a			
v	v'		z		ž						
m	m'		n	ń							
		r	l	l̇							
					y						

Some dialects have two more vowel phonemes, /ɛ/ and /ɔ/.

Some dialects have phonemic stress and pitch; the southern dialects have non-phonemic stress on the first syllable.

Voiced stops do not occur finally.

LOWER LUSATIAN

For the Lower Lusatian dialect spoken by Protestants at Kottbus on the Spree, Harold L. Klagstad gives the following phonemic inventory:

p	p'	t	c	ć	č	k	
b	b'	d		ź	ž	g	
f	f'		s	ś	š	x	h
v	v'		z	ź	ž		
m	m'		n	ń			
			r	ŕ			
			l	ĺ			
					y		

i	ə	u
e		o
	a	

Stress occurs on initial syllables.

Voiced as well as voiceless stops occur finally.

CZECH

In addition to the general sources, the following source was consulted for Czech: Robert R. Leed, Historical Phonology of Czech, Ph. D. Dissertation, Cornell University (1958). The phonemic inventory for Czech is:

p t c ť č k

b d ʒ ď ǯ g

f s š x

v z ž γ

m n ń

l

r ř

y

i u

e o

a

plus length

plus syllabic /ḷ ṛ/

Some sources exclude /ʒ ǯ g f/, as they occur only in loanwords.

Long /e· o·/ contrast only in loanwords.

SLOVAK

The following sources were consulted for Slovak (in addition to the general sources above): B. Hála, Zakady Spăsouné vý Slovnosh Slovensiu (Prague, 1929) and W. K. Matthews, Slouvak, Le Maître Phonetique 91.1-3 (1949).

The following consonant inventory represents a general consensus:

p t c č ť k

b d ʒ ǯ ď g

f s š x

v z ž γ

m n ń

l ĺ

r

y

Bidwell says that many speakers replace /ĺ/ with /l/.

/ć/ and /ʒ́/ do not occur in some dialects.

For short versus long vowels most sources indicate the following coexistent systems:

i	u	i·		u·
e	o	e·		o·
æ	a		a·	

plus long versus short syllabic /l̥/ and /r̥/

Bidwell indicates that /æ/ is replaced by /e/ by many speakers, yielding the same five-vowel system for short as for long vowels. Stankiewicz agrees for East Slovak, but for the Standard gives a 3(F B) plus length system, i.e., all six vowels occur both short and long.

BULGARIAN

For Standard Literary Bulgarian and some dialects the following sources were consulted (in addition to the general sources listed above): H. L. Klagstad, A Phonemic Analysis of Some Bulgarian Dialects in American Contributions to the Fourth International Congress of Slavicists (1958) and The Phonemic System of Colloquial Standard Bulgarian, SEEJ XVI. 42-54 (1958); C. T. Hodge, Bulgarian Basic Course, FSI, Washington D. C. (1951); H. I Aronson, Morphophonemic Patterns of the Bulgarian Inflections, Ph.D. thesis, Indiana University (1961); and L. Ardrejčin, N. Kostov, and E. Nkolov, Bəlgarska Grammatika, (Sofia, 1947).

p	p'	t	t'	c	c'	č	k	k'
b	b'	d	d'	ʒ	ʒ'	ǯ	g	g'
f	f'			s	s'	š	x	
v	v'			z	z'	ž		
m	m'			n	n'			
				l	l'			
				r	r'			
						y		

i ə u　　í ə́ ú

é á ó

Hodge and Bidwell treat the palatalized consonants not as separate unit phonemes (as given in the inventory above) but as clusters of consonants plus /y/, which occur only before front vowels. There is more agreement among the sources in the treatment of /ĺ ń ḱ ǵ/ as palatalized phonemes than in the treatment of the other palatalized phonemes. The palatalized consonants shown in the inventory above are given in the inventory of the Standard dialect by Andrečin, Lunt, Aronson and Klagstad. Klagstad analyzed the Sophia Colloquial Standard, a compromise between the Colloquial Standard of Ternovo and the regional dialect of Sophia (the Sophia local dialect does not have phonemic palatalized consonants, but the Ternovo Colloquial does).

/ʒ/ occurs in some dialects, but is not generally considered a phoneme of the Literary Standard. Stankiewicz considers both /ʒ/ and /ǯ/ as marginal (with East Bulgarian not having even a marginal /ǯ/).

Unstressed /e/ and /o/ are optionally dropped in the Literary language, occurring only in formal style.

For the Sophia dialect, Klagstad indicates a 2(FCB) six-vowel system with an SGC of stress; i.e., all vowels occur both stressed and unstressed.

Klagstad (1958) gives the following inventories for other Bulgarian dialects:

Gabare Dialect

p	t	c	ć	č	k
b	d		ʒ́	ǯ	g
f		s		š	x
v		z		ž	
m	n		ń		
	l		ĺ		
	r				
			y		

i	ə	u
e	a	o

plus stress

Lovec Dialect

p	p'	t	t'	c	č	k	k'
b	b'	d	d'		ǯ	g	g'
f	f'	s	s'		š	x	
v	v'	z	z'		ž		
m	m'	n	n'				
		l	l'				
		r	r'				
					y		

i	ə	u	í	ə́	ú
			é	ɑ́	ó

Nova Nadezda Dialect

p	p'	t	c	c´	č	k
b	b'	d		ʒ́		g
f	f'	s			š	x
v	v'	z			ž	
m	m'	n		n´		
				l´		
				r´		
				ṿ		

i	ə	u
e		o
plus stress		

Klagstad also gives the vowel inventories for a number of dialects. The Erkeč dialect has a 3(FB) vowel type; the Boboševo has a 2(FB) over neutral vowel type; and the Rhodopian Momčilovci dialect has a 3(FB) over neutral vowel type; all occur with stress:

i u	i u	i u
e o	e o	e o
æ a	a	ɛ ɔ
		a
Erkeč	Boboševo	Rhodopian Momčilovci
3(FB)	$\frac{2(FB)}{N}$	$\frac{3(FB)}{N}$

MACEDONIAN

For Macedonian the following sources were consulted, in addition to the general sources above: Horace C. Lunt, Grammar of the Macedonian Language (Skopje, 1952) and review of Makedonskagramatika Krumekepeski (Skopje, 1950); W. K. Matthews, The Cyrillo-Methodian Source of Modern Macedonian Phonology, Canadian Slavic Papers III. 1-6 (1958); and Harold Klagstad. The Literary Standard is based on the central dialects (west as far as the Vardar River), and makes the following phonemic distinctions:

p	t	c	ć	č	k	k'				
b	d	z̨	ź̨	ž̨	g	g'				
f	s			š	x		h	i	ə	u
v	z			ž				e	a	o
m	n		ń							
	l		ĺ							
	r									
			y							

/ə/ occurs in various dialects, in unassimilated foreign loans, and before /r/ plus a consonant, e.g.[ə r š] rye, where[ə] plus[r] are alternatively analyzable as syllabic /r̥/.

/ć ź̨/ are dorso-velar, or frontal-palatal stops with fricative offglides, but less fricative offglide than Serbo-Croatian /ć ź̨/.

/v/ does not occur in some dialects.

/y/ is rare in some dialects.

/x/ occurs only in loan words

/h/ has a very limited distribution.

Stress is fixed in Literary Macedonian, but in most of the eastern dialects (primarily in Bulgaria) it is phonemic, as in Bulgarian.

Stankiewicz indicates that palatalized labials, dentals and velars (plus /r' l'/) are phonemic in Eastern Macedonia.

SERBO-CROATIAN

The following sources, in addition to the general references above, were consulted: Roman Jakobson, On the Identification of Phonemic Entities, TCLC 5.205-213 (1949), Carleton T. Hodge, Serbo-Croatian Phonemes, Lg. 22.112-20 (1946), Helmut von Lüdtke, Bonn, Das Prosodische System des Urslavschen und seine Weiterentwicklung im Serbokroatischen, Phonetica 4.125-156 (1959), and Paule Ivic, Die Serbokroatischen Dialekte.

The consonant inventory for Standard Literary Serbo-Croatian is given below:

p	t	c	ć	č	k
b	d		ʒ́	ǯ	g
f		s		š	x
v		z		ž	
m	n		ń		
	l		ĺ		
	r				
			y		

Most dialects have no /ʐ/ (see below).

/ń/ and /ĺ/ are sometimes treated as clusters /ny/ and /ly/.

Jakobson indicated a 2(FB) over neutral vowel type, plus syllabic /r̥/, combinable with length; both short and long vowels are combinable with rising pitch:

i u

e o r̥

a

plus length; plus rising pitch.

Klagstad points out that syllabicity of [r] is predictable except in one word: / umr̥o / *he died.*

Bidwell gives the following description of length and pitch in Serbo-Croatian (incorporating Hodge's solution):

"All words in Serbo-Croatian, except for a small number of enclitics and proclitics which do not occur independently, have a lexical stress, which may occur on any syllable but the last, in words other than recent borrowings. Under primary stress the presence or absence of "rising tone" may be detected (under secondary stress the tone contrast is neutralized in most standard speech). Phonetically, this consists in carrying over increased stress into the following syllable and, in the case of long vowels, the onset of stress in the second half thereof. With the contrasting "falling tone", stress is never carried over into the next syllable and in long vowels the onset is at the beginning of the vowel. If one accepts Hodge's analysis of long vowels as double vowels, "rising tone" may be analysed as a double stress extending over two syllables (in some cases a single stress on the second half of a long vowel), while "falling tone" is a single stress, on the first half if the vowel is long. If one regards long vowels as units however, it is necessary to posit at least one additional phoneme, namely a phoneme of rising tone. In any event, the presence or absence of the rising feature combined with the presence or absence of vowel length comprise the four traditional "accents" of SCr."

The following consonant phonemes are given by Ivic for the Stokavian dialects.

East Hercegovina

p	t	c	č	ć	k
b	d		ž̨	ź̨	g
		s	š	x̭	
v		z	ž	γ̭	
m	n		ń		
	l				
	r				
			y		

Among Moslem speakers /ć ź̨ x̭ γ̭ / are lacking and /h/ and /z̨/ are added (but the latter only in Dubrovnik).

The Sumadija-Vojvodina dialect differs only from the above in lacking /x̭ γ̭/, except in Banat where /ć ź̨/ are also lacking, while there are the additional phonemes /f/ and /h/. Late Ikavian has the same phonemic inventory as Sumadija-Vojuodina, with the Catholic speakers lacking only

/x̯ γ̯/, and the Moslem speakers lacking /ć ʒ́ x̯́/ and adding /f/ and /h/. In Karlobag and Šibenik (Late Ikavian) /ć ʒ́/ replaces /č ǯ/, and /š ž/ are missing (as well as /h/, and probably /f/). In Mostar (Late Ikavian) /ń/ is missing as well.

In Zeta-Lovćen (Stokavian) the phonemes /ć ʒ́ x̯ γ̯/ are present as well as /f/ and /h/. Otherwise the inventory is the same as that for East Hercegovina, except that /ʒ/ appears in some localities and some Moslems merge /č/ and /ć/.

Kosovo-Rasava has /ć ʒ́/ but not /x̯ γ̯/. Moslem speakers also have /f h/, but not Catholics.

In Istric Ikavian most, but not all, localities have /ć ʒ́/ and /f h/. In the extreme northwest /č/ and /š ž/ are missing. The phonemes /ʒ ǯ/ do not occur.

The 'Rumanian dialects with /I/' all have /ć ʒ́/ and /f h/. The Krašovani and Rekaš dialects both have /č ǯ š ž/, all of which are missing in Gallipoli.

In Slavonic the maximum phonemic system is like that of East Hercegovina, but some speakers do not have /ć ʒ́/, and others do not have /č ǯ x γ/.

SLOVENE

The following sources were consulted for Slovene: I. Lehiste, The Phonemes of Slovene, IJSLP 4.48-66 (1961); Slavo Klemenčic, Slovene Literary Pronunciation, Le Maître Phonetique No. 111.3-5 (1959); Jože

Toporišič, Probleme der Slovenischen Schriftsprache, Scando-Slavica 6.53-74 (1956); Edward Stankiewicz, The Vocalic System of Modern Standard Slovenian, IJSLP 1.70-76 (1959).

p	t	c	ć	č	k
b	d				g
f	s			š	x
v	z			ž	
m	n		ń		
	l		ĺ		
	r				
				y	

Some investigators, as I. Lehiste and Slavo Klemenčic indicate a /w/ phoneme, which contrasts with /v/ only in initial position before /l/, e.g. / vléci / to pull, / wlíti / to pour in.

Lunt indicates that /ń ĺ / phonemes occur, but only in artificial speech. Some dialects have only /č š ž ń/ instead of /č ć š ž ń ĺ /.

Stankiewicz indicates the following coexistent vowel systems for Modern Standard Slovenian, showing 21 vocalic phonemes. The important distinction is pitch, as length is concomitant with rising and falling pitch. Stress occurs on words with length (rising or falling pitch) on the long syllable, while in words without length the stress falls on the final syllable. The vocalic phoneme /r/ is combinable with /´/ and /`/.

í	ú	ì	ù	i	ə	u		
é	ó	è	ò	e	o	a	plus	r̥
ɛ́	ɔ́	à						
á								
rising long		falling long		short				

Prescriptive grammarians (R. Tesniere is cited) have recognized an innovating dialect which has 19 vocalic phonemes (plus /r̥/), of which 7 are long and 12 are short (short stressed and short unstressed):

i	ə	u	i·	u·
e	a	o	e·	o·
plus stress			ɛ·	ɔ·
			a·	

Stankiewicz, however, rejects this solution and proposes the following (with a single stressed vowel):

i	ə	u	i·	u·		
e	a	o	e·	o·	ə́	r
short			ɛ·	ɔ·		
			a.			
			long			

Stankiewicz also mentions a new system of oppositions gaining ground among the younger educated speakers of Ljubljana, called by Tesniere the colloquial form of the literary language, having 8 stressed and 6 unstressed vowels with neutralized length.

í ú   i ə u
é ə́ ó   e a o
ɛ́ á ɛ́

I. Lehiste and Lunt add falling long /ɛ̀ ɔ̀/ to the inventory given by Stankiewicz for Modern Standard Slovenian vowels. Slavo Klemenčic further adds /I U ɛ ɔ/ to the short vowel system. Jože Toporišič does not include /ɛ ɔ /, but includes long falling and long rising /ə /:

i ə u   í ə́ ú   ì ə̀ ù
e a o   é ó á   è à ò

W. K. Matthews does not admit tone as distinctive in the literary language, but reports that three tones still survive in the dialects (dialect information is available in F. Ramovš. Dialecktološka karta slovenskoga jazika, Ljubljana, 1931).

THE BALTIC BRANCH OF INDO-EUROPEAN

6. The two extant representatives of Baltic, Lithuanian and Lettish (Latvian) are spoken by about 5 million people primarily in Lithuania and Latvia, respectively, where they are also the official national languages. Of the two Lithuanian has the greater number of speakers (by one million). A great number of Latvians and Lithuanians were displaced during World War II and the Russian occupation, and now live in Central and Eastern Europe, and North and South America.

Despite the late date at which Baltic literature appeared, the Baltic languages are recognized as the most conservative of the Indo-European family, containing many features already lost in the most ancient Greek, Sanskrit and Latin records. Among these features Baltic has (1) retained the Indo-European short and long vowels, with the exception of Proto-Indo-European */ə/, */a/, and */o/, which have merged into /a/; (2) pitch accent; (3) seven of the eight Proto Indo-European cases; and other phonological and morphological retentions. An example of the sometimes striking similarity to reconstructed Indo-European is the word for son, sūnùs (Lithuanian), which is identical to the reconstructed Indo-European *sūnus. Other Indo-European languages show new developments, as Sanskrit sūnúh, Gothic sunus, Old English sunu, and Old Church Slavonic synŭ.

Baltic is characterized by its large number of case forms, seven of which are retained from Proto-Indo-European. The eight cases distinguished for Latvian are: nominative (miẽst-as *city*), genitive (miẽst-o), dative (miẽst-ui), accusative (miẽst-a), vocative (miẽst-e), instrumental (miest-ù), locative (miest-è), and illative (miẽst-an, *into the city*). There are also remnants of a 'directive case' and an 'adessive case', both of which occurred regularly in 16th and 17th century Baltic literature.

The early dialectal differentiation of Baltic led to a split between East Baltic and West Baltic, perhaps not later than 500-300 B.C. (although placed earlier by some sources). The Western Baltic group consists of old Prussian and the closely related dialect of Suduvian, or Jatvigian, both of which became extinct as a result of early wars with the Germans. The last old Prussian dialects were spoken in the 17th century (though traces of Jatvingian were found as late as the 19th century). Other extinct languages or dialects are: *Curonian (Kurish), which formed a connecting link between West and East Baltic (extinct since about 1600); *Zemgalian (SemiGallian), spoken until the late Middle Ages in the Musalie lupe River basin (extinct since 1450); *Selonian, spoken in Middle Ages near Dvinsk-Daugavpils (extinct since 1400).

Curonian, Zemgalian, and Selonian were absorbed by Latvian and Lithuanian and have left only meager records.

The only Baltic languages still spoken are Lithuanian and Lettish, these are both East Baltic languages. East Baltic remained rather homogeneous for a long period after the split between East and West Baltic, until about 400-600 A.D. when Curonian split. The other languages, Lettish, Zemgalian and Selonian slowly differentiated over a period from 700 to 1100 A.D.

The following Baltic languages are the best known:

(1) *Old Prussian, spoken in Prussia, or Borussia (in an area later called East Prussia) between the Vistula and Niemunas (Memel) Rivers. It is known from the 15th and 16th centuries through three Protestant catechisms, two short lists of words; and through a few personal and place names.

(2) Lithuanian is spoken today by about 3 million people primarily in Lithuania, where it became the official language in 1918. There are Lithuanian speakers found in the United States (400,000), Brazil (40,000), Argentina (35,000), Great Britain (12,000), Uruguay (10,000), Canada (10,000), Central Western Europe (70,000), and Siberia.

The two main dialects of modern Lithuanian are:
Shamaitish (Samogitian, Žemaitish, Žemaičiai, Low Lithuanian);
Aukshtaitish (Aukstaiciai, High Lithuanian).

The Modern Standard Language is based on a sub-dialect of Aukshtaitish. The earliest written Lithuanian document was a Protestant catechism in the Shamaitish dialect.

(3) Lettish (Latvian) is spoken today by about 2 million people, primarily in Latvia. The merging of the Latvian tribes -- such as the Latgallians Zemgallian, Sels -- and assimilation with such neighboring peoples as West Finnish Livonians, resulted in three main dialects:
Eastern (High, Upper) Lettish, including Lagalian;
Central (Middle) Lettish;
Tamian (Western) Lettish.

The accepted standard language, written in a Roman alphabet, is based on Central Latvian. The earliest literary document was a catechism published in 1585.

Internally there are a number of different phonological reflexes distinguishing West and East Baltic. In Old Prussian the Proto-Indo-European /ā ō/ fell together, but remained separate for the most part, in Eastern Baltic. Proto-Baltic /ty dy/ became dental /t d/ in West Baltic, became affricate /č ǯ/ in Lithuanian, and became fricative /š ž/ in Lettish. There are also, however, a number of reflexes which cut across the East-West division.

There are also a number of features separating modern Lettish and Lithuanian: (1) Lithuanian /š ž/ corresponds to Lettish /s z/; (2) the Lithuanian affricates /č ǯ/ correspond to Lettish /š ž/ (3) Lithuanian /an, en, in, un/ corresponds to the Lettish development of /uo, ie, ī, ū/ and (4) stress is fixed (predictable) in Lettish, but phonemic in Lithuanian.

The typical modern phonemic patterning of Baltic is suggested by the following charts after Eric Hamp (Buividze Lithuanian Phonemes, IJSLP I, 1959) and Valdis J. Zeps (Latvian and Finnic Linguistic Convergences, Indiana University Publications, Uralic and Altaic Series, Vol. 9, 1962):

LATVIAN

p	t	c	č	ќ	k	
b	d	ʒ	ǯ	ǵ	g	
f		s	š		(x)	(h)
v		z	ž			
m		n		ń		
		l		ľ		
		r		ŕ		
				y		

Items enclosed in parentheses occur only in loans.

The 'Latgalian Standard' lacks the phonemes /ќ ǵ ń ľ ŕ/.

Zeps sets up phonemically long nasals, laterals, /r/ and semivowels in the Latvian Standard, /m̄ n̄ l̄ ļ̄ r̄ v̄ ȳ/, but not for the Latgalian Standard.

Zeps discusses at length the vowel systems in Latvia, but the Latvian Standard seems to be:

i	u
e	o
æ	a

plus length /·/

plus falling accent /ˋ/

plus rising accent (long vowels only) /~/

The Latgalian Standard, however, has the following system:

i		u	i·	u·
e	ə	o		
æ		a		a·

plus glottalization /ʔ/

LITHUANIAN

p	t	č	k	
b	d	ǯ	g	
f	s	š	(x)	(h)
v	z	ž		
m	n			
	l			
	r			
		y		

i	u
e	(o)
a	

plus length

Items in parentheses occur only in foreign loans.

Some dialects have a phonemic contrast of /l/:/ł/.

Hamp analyzes the accent system as consisting of primary stress /ˈ/, secondary stress /ˌ/, and 'peak' tone or pitch /~/.

GREEK

7. Modern Greek (Romaic, Neoellineki, Neo-Hellenic, Grec, Neugriechisch,Graecae, etc.) is spoken principally in the Kingdom of Greece (Vasilion tis Elladhos), a country of southern Europe which comprises the southern peninsular projection of the Balkans, the northern foreshore of the Aegean as far east as the Maritsa (Evros) River and, except for Imroz and Bozca Odo (Tenedos), all the main islands of the Aegean, including Crete, the Dodecanese, and the Ionian Islands off the west coast of Greece in the Ionian Sea. It is bounded on the north by the nations of Albania, Yugoslavia and Bulgaria and on the east by Turkey. It is bounded by seas to the west and south—the Ionian and Mediterranean Seas, respectively.

In Greece 95.6 percent (7,300,000 persons) speak modern Greek as their mother tongue. The remaining 4.4 percent of the population either speak Turkish or another Indo-European language—one of a half dozen others:
Turkish with 180,000 speakers consisting of a Moslem minority of 92,000 in Western Thrace and the Dodocanese Islands, in addition to 88,000 Turkish speaking Greek immigrants from Asia Minor;
Macedonian Slavic with 41,000 speakers located mostly in western Macedonia;
Vlach, a dialect of Rumanian, spoken by 40,000 (Britannica) or 60,000 (G. Bonfante) Koutsovlachs;
Albanian spoken by 23,000 (Britannica) to 50,000(G. Bonfante) people in parts of Attica, Argolis, and the islands;

Pomoh, a dialect of Bulgarian, spoken by 19,000 Moslems in the Rhodope Mountains of Thrace;

Tsakonian Greek spoken by more than 10,000 speakers on eastern Peloponnesian coast of the Arcadian department;

Armenian refugee speakers;

Russian refugee speakers.

The Albanian and Vlach-speaking populations are mainly bilingual, with Greek as the second language.

In addition to the Greek speakers of the Kingdom of Greece, there are Greek speaking communities in other Mediterranean and Balkan countries, as in:

Turkey, especially in Istanbul, there are 95,000 persons who speak modern Greek;

Alexandria Egypt (60,000 speakers);

Cyprus, (425,000 Greek speakers constitute 80 percent of the population of the island);

Italy (two colonies of more than 30,000 people in the two southern peninsulas of Italy);

Balkans, (many Greek speakers are found in colonies in southern Albania, Eastern Yugoslavia, Rumania, Bulgaria, and the U.S.S.R. states surrounding the Black Sea, especially Crimea, Moldavia, and Georgia).

Greek speakers are also found in most of the Western European countries, Africa, and especially in various areas of North and South

America due to recent immigration dating from World War II.

Although Modern Greek, a branch of the Indo-European language, is closely related by virtue of shared innovations with Italic and Indo-Iranian branches, Giuliano Bonfante demonstrates its even closer relationship to Armenian by the following phonological innovations: (1) development of vowel before word-initial r-, l-, m-, n-; (2) replacement of word-initial and intervocalic /s/ by /h/ and later loss of the /h/ in these environments; (3) loss of intervocalic /y/; (4) the replacement of syllabic l, r, m, n by the sequences al, ar, am, an respectively; (5) the replacement of ly, ry, my, ny following a vowel, by a vowel + /y/ sequence preceding the l, r, m, and n respectively (i.e. VCy → VyC); (6) loss of /w/ from labialized velar stops following /u/; (7) replacement of labialized velar stops before /i/ and /e/ by a dental stop (in Greek) or alveo-palatal affricate (in Armenian); and (8) early loss of word final /t/ and /d/.

Carl D. Buck, in his Comparative Grammar of Greek and Latin, sets forth distinctive features common to all the old Greek dialects, from which is to be inferred a period of common development, a relatively unified Greek language. The Greek dialects all maintained the old vowel system of /i e a o u/ plus a long-short distinction and the pitch accent: high /´/ (acute) versus low /`/ (grave), in addition to a sequence of high-low (in that order) occurring only on long vowels (the circumflex accent). The Indo-European diphthongs are also maintained, /ei, ai, oi, ui, eu, au, ou/ for short vowels and /ēi, āi, ōi, ēu, āu, ōu/ for the long vowels.

Myceaean Greek (Linear B) gives evidence of the retention of the Indo-European labialized velar stops, /k^w g^w k^{hw}/, thus giving the following consonantal stop system:

p	t	k	k^w
b	d	g	g^w
p^h	t^h	k^h	k^{hw}

The classical dialects, however, show stops having only three linear distinctions:

p	t	k
b	d	g
p^h	t^h	k^h

The reflexes of the labialized velars /k^w g^w k^{hw}/ occur respectively as /p b ph/ before /a o/, as /t d th/ before /i e/, and as /k g k^h/ before or after /u/. Word-finally all consonants but /r s and n/ were lost.

The phonology of standard spoken modern Greek has a voiced and voiceless series of stops; and of fricatives. The voiceless and voiced fricatives are reflexes, generally, of the voiceless aspirates and voiced stops respectively; but the voiced stops derive from a split of /p t k/ following nasals, when the nasals were lost in that pre-stop position or when contrasts were introduced by loanwords having a sequence of a nasal plus voiceless stop. The modern Greek vowel system, like that of ancient Greek is of the 2(FB) over N type—/i e a o u/—but in contrast to ancient Greek modern Greek has no phonemic length. The modern Greek accent of stress corresponds to the older pitch accent. The phonology of modern Greek as described by Andreas Koutsoudos (IJAL, 1962) is as follows:

p		t	k	i	u
b		d	g	e	o
f	θ	s	x	a	
v	δ	z	γ	(plus stress)	
m	n				
	l				
	r				
			y		

The /y/ consonant may occur following any consonant prevocalically, in which environment, in the case of the /k g x γ /, there simultaneously occurs palatal allophones of the /k g x γ/, viz. [k^y, g^y, ç, y].

The spoken Greek of today is the latest development of the Attic dialect of ancient Greek, which, out of many dialects, became predominant. From Attic developed the Koine, or common dialect, which spread over Alexander's empire and which displaced all of the classical dialects except Tsakonian. Simultaneous to the spread and development of the Koine, an Atticistic rhetorical-literary movement, in reaction to the innovations of Koine, arose attempting to revive classical Attic grammar and vocabulary. The official language of the Medieval or Byzantine period, an heir to the Atticistic tradition, also represents an attempt to imitate the classical Attic models.

This divergency between the literary and spoken language is perpetuated in the language problem of Modern Greek: Katharevousa versus Dhimotiki. The use of two sharply differentiated linguistic varieties: the learned archaic literary dialect (Katharevousa) and the colloquial based literary dialect (Dhimotiki) has been a striking feature of modern Greek ever since the medieval period. In addition to these two varieties there is an intermediary conversational dialect, which constitutes a third entity; it is the standard spoken modern Greek (mixed, colloquial Dhimotiki). The situation created by the separate existence of the Katharevousa and the Dhimotiki has frequently become an emotional political issue, sometimes marked by strife and bloodshed. The origin of the two separate varieties in modern Greek is concisely set forth by F. W. Householder with assistance by Costas Kazazis (Studies in Modern Greek, III: Greek Triglossia): "[In] creating an official written or literary language for [the new Greek nation], ... at first ... many of the responsible authorities took the line that the language had not changed since New Testament times, and that all the apparent changes were simply ignorant, slovenly speech which could be cured by education; the only official grammar was the grammar of Ancient Greek. Of course no one ever succeeded in writing pure Classical (or even Koine) Greek, and few people even tried to go that far. The basis usually taken was rather late medieval and early modern scholastic-legal-theological Greek, a fairly homogeneous development of the Byzantine literary language, but remote in many ways from the language spoken by anyone (even an archbishop). This was the earliest form of Katharevousa, and over the years it has been modified successively many times into a closer and closer agreement with the spoken language, from which, however, it still differs markedly in certain sacrosanct details of spelling, morphology, and vocabulary (and a few minor points of syntax.)

"The other literary language, Dhimotiki, was based more or less closely on the spoken language of the 19th century, also with several centuries of similar writing as a guide. It was the great achievement of Manolis Triandaphyllidhis (and his associates) to provide in 1941 an official grammar of this language, which could be used to settle uncertainties and disputes over spelling, inflection and derivation. The standard practice of today differs on hardly any point from the recommendations of Triandaphyllidhis, There is even now no equivalent official grammar of Katharevousa."

In function, Katharevousa serves as the official tongue of the Greek state. It is spoken only on formal occasions such as in Greek Orthodox liturgy, in the Greek parliament, and in academic lecture, but it is used almost exclusively for official, and scientific writings as well as by most newspapers. The function of the Dhimotiki is more general in that its sphere is in those areas not designated for Katharevousa. Nevertheless, some proponents of Dhimotiki would have it used exclusively for all functions.

The standard spoken modern Greek is based on the dialect of the region around Corinth, the first national capital of the modern Kingdom of Greece; however, in addition to the standard spoken language, modern Greek has considerable dialectal variation which is geographically correlated. There are two major phonological isoglosses by which the modern Greek local dialects fall into four major groups. The first isogloss, suggested by Manolis Triandaphyllidhis, running north and south, divides the dialects into the eastern and western groups. The eastern dialects comprise those local varieties spoken on Chios, the Dodecaneses, Cyprus, and in Asia Minor (before the Turkish-Greek exchange in 1922-23). The western group comprises the remaining geographical area of Greece. The Eastern dialects retain a word final /n/, whereas the western dialects do not.

The second isogloss, outlined by Paul Kretschmer in 1905, runs east and west, dividing the dialects into a north and south group. The northern group comprises: continental Greece as far south as the Gulf of Corinth and the northern boundary of the Peloponnese; the two northern Ionian Island groups of Corfu (Kerkira and Paxoi), and Leukas (Levkas); and the Aegean Islands, which are, generally, north of a line running east from the southern tip of Attia to Asia Minor, including the Andros Island of the Cyclades plus Tinos (which is south of the line) and Samos (but not Ikaria) of the Dodecaneses. One exception in this geographic group is the island of Chios, deep in the northern dialect territory, which must be grouped with the southern dialects.

The northern dialects share the unstressed vowel system of /i a u/, whereas in the southern dialects the unstressed vowels /i e a o u/ correspond to the stressed ones /í é á ó ú/. The reduced three vowel system of the northern dialects resulted historically from two innovations: the loss of /i/ and /u/ in unstressed environment; and a subsequent phonetic shift of the e and o to i and u respectively. As a consequence of the intersection of the two isoglosses the modern Greek dialects fall into four geographically correlated groups: Northwestern group, Northeastern group, Southwestern group, and Southeastern group. It is obvious that since the Northwestern and Southwestern groups encompass the greatest land area, the majority of modern Greeks speak dialects of these two groups. Other phonological isoglosses (as well as morphological) cut across the above mentioned dialect groups, but are less extensive geographically.

The great diversity in modern Greek from region to region is illustrated by the dialect list below. Where available, the characteristic divergency of the dialect from standard modern Greek is discussed.

Saracatsan

Maniote, including:

    Northern Maniote

    Southern Maniote

Epire-Rumel

Cypriot

Cretan

Ionian Island

Northern Aegean Island, including:

        - Lesbian
        - Samotic
        - Euboean
- Asia Minor group, including:
    - Pontic
    - Cappadocian
- Southern Italian Greek
- Thessalian-Macedonian-Thrace
- Stera Hellas
- Peloponnese (not including Maniote)
- Cyclades Island
- Dodecanese Island.

Saracatsan (Tsarakatsanai, Sarakachani, Karakachani) is spoken by the Saracatsans, a tribe of Greek nomads living in the central mountainous regions of continental Greece in the divisions of Epire, Thessaly, Macedonia, and Thrace. There are also reports of them being situated across the national Greecian borders in the mountains of Albania, Yugoslavia, and possibly Bulgaria. During the summer they graze their sheep herds in the mountains, but descend to the valleys for the winter months. In 1925 there were an estimated 6,000 Saracatsans in the Epire Division.

For the groups of Saracatsans living in the Thessaly, Macedonia and Thrace Divisions, no population statistics were available. Formerly, the Saracatsans often crossed the national boundary into Albania, Serbian

Yugoslavia, and Bulgaria, but the political situation now complicates such a passage. In 1910 Ischirkoff (cited by Carston Hoeg in Saracatsans: Une Tribe Nomade Grecque, 1925), reported 4,600 Saracatsans living in Bulgaria.

Although modern Greek and Saracatsan are mutually intelligible, Hoeg notes that 'it is difficult . . . for an Athenian to understand a Saracatsan [translated]'. Phonologically there are many points in which Saracatsan differs from standard spoken modern Greek, but the most significant is its additional set of palatal affricates and fricatives (/č ǯ š ž/). Saracatsan is most closely related to the Rumel-Epire dialect of modern Greek.

Maniote is the Modern Greek dialect spoken on Magna, the central one of the three southern peninsular extensions of the Peloponnese. André Mirambel (Etude Descriptive du Parler Maniote Méridional) reports in 1929 that the population of this region was over 47,000 people, concentrated in 271 towns and villages. The significant phonological features, which characterized this dialect in contrast to standard modern Greek are primarily differences in phoneme distributions: (1) 'the loss of intervocalic /v, δ, γ, l/; (2) the occurrence of /u/ medially and finally as a reflex of Classical Greek /ō / for which modern Greek has /o/; (3) the occurrence of /u/ or /iu/ as a reflex of Classical Greek /u/ for which modern Greek has /i/.

The Maniote Greek spoken in the towns and villages along the western coast is more homogeneous than that spoken in the moutainous regions. Maniote Greek has two subdialects, Northern Maniote and

Southern Maniote. The southern dialect has for the /k γ x/ of standard modern Greek the characteristic reflexes [ts, š, ž] before /i/ and /e/ which correspond to /k^y, γ̯, x̯/ of the northern dialect. The southern dialect extends as far north as a line extending southeast from Areopolis on the west coast to Kotrones on the east coast. The Northern Maniote dialect is spoken in the remainder of the Magna region north of the line.

The Epire-Rumel dialect is the dialect of Modern Greek spoken in Epire Division and in the region of both Thessally and Stera Hellas which lies approximately between and around Karpension and Kardhitsa in the Pindus mountain range. The phonological characteristics of the Greek of this area are: (1) loss of unstressed /i/ and /u/; (2) occurrence of consonant clusters (as a result of #1) which do not occur in modern Greek e. g. word initial /psl/, /mpl/ and /xst/; (3) the loss of /r/ between a velar /x/ or /k/ and a following /st/; (4) the occurrence of stress on the fourth syllable from the end, which can occur no further from the end of a word than the third syllable in standard modern Greek; (5) loss of /γ/ intervocalically; (6) the occurrence of word-final /n/ only utterance finally.

The Cypriot dialect of Modern Greek is spoken by 425,000 persons who make up 80 percent of the population of Cyprus. It is spoken throughout Cyprus except for those regions inhabited exclusively by Moslems (where Turkish is spoken), and in the village of Kormakiti in the northwest (where the inhabitants speak Arabic). The Cypriot dialect is mutually intelligible with the other modern Greek dialects but is most similar to the dialect spoken on Rhodes. Some phonological characteristics of Cypriot are: (1) the loss of /β δ γ/ intervocalically; (2) the replacement of /gy/ after a consonant (except /r/) by /ky/ and after /r/ by /k/; (3) the replacement of /δ/ after /v/ or /γ/ by /t/; (4) the occurrence of palatalized allophones of /k γ x s/ (viz [č, y, š, š] respectively) before /i/ and /e/; (5) replacement of /x/ after /r/ by /k/.

The Cretan dialect of modern Greek is spoken by approximately 6 percent of the total Greek population (438,000 people) living on the Island of Crete. Some of the phonological characteristics of Cretan which distinguish it from standard modern Greek are: (1) the occurrence of alveopalatal allophones [č, ž, š, ǰ] of /k, γ, x, and g/ respectively before /i/ and /e/; (2) the loss of nasals before /b, d, g, θ/ word medially; (3) the replacement of /z/ by /r/ before nasals; (4) the replacement of /l/ by /r/ before /p t k m/ and even intervocalically in the region of Sphakia; (5) loss of unstressed /i/ after /s r l/; (6) loss of final /r/; (7) the replacement of /t/ by /θ/ before /i/ plus vowel.

The Ionian Island dialect of modern Greek is spoken by approximately 200,000 inhabitants of the Ionian Island off the west coast of mainland Greece. This dialect has many Italian borrowings in its lexicon. The phonological characteristic of this dialect is the loss of nasals before /b d g/.

The dialects of the Aegean Islands, included in the northwest dialect group (estimates of speakers ranges from 200,000 to 400,000 persons), in the islands of the Lesbos, Samos, Euboean, and Magnesian departments, are differentiated from the northwestern continental dialects by the following phonological features set forth by Paul Kretschun (Der Heutige Lesbische Dialekt, 1905): (1) the preservation of unstressed /i/ and /u/ in the environment preceding /r/; (2) the loss of nasals before stops and fricatives (which is valid also for many of the southern islands). Lesbian, one of the subdialects of this Northern Aegean Island group, is spoken by over 100,000 people on the islands of the Lesbos department. It is characterized by the replacement of /t/ and /d/ before /i/ by /k/ and /g/—that is, by [k^y] and [g^y], respectively.

Some dialects of modern Greek spoken in Asia Minor in the 19th century are today almost non-existent. The bulk of the speakers of these dialects immigrated from Turkey in 1923. The people settled in different dialect areas of the Kingdom of Greece and are being steadily integrated into the respective dialects of the area. K. Hatzikakis (Einleitung in die neugriechische Grammatik, 1888) groups the dialects of Asia Minor into the highland dialects (which share the /i a u/ unstressed vowel system with the north Greek dialects) and the coastal dialects (which are southern dialects). Archelaos (1899) states that northern Greek dialects of Asia Minor resulted historically from dispersed enclaves of earlier Lycaonian and Cappadocian (including Pontic) dialect speakers. The dialect today in Livisi of the old Lycian area is representative of the southern dialect in that it has /i e a o u/ unstressed vowel system. Two of the pre-contemporary modern Greek dialects of Asia Minor are discussed below.

The Pontic dialect of modern Greek was spoken in 800 villages in Turkey along the eastern section of the southern Black Sea coast. Although most of the Greek speakers of this dialect were removed to Greece in 1922 and 1923, the Pontic dialect is still spoken by the few Greeks who remained in Turkey in the Pontic region as well as by the Greek-speaking Turks of that region. There are three subdialects of Pontika. Phonologically the Pontic dialect is characterized thus: (1) the occurrence of /e/ in contrast to standard modern Greek's /i/ as the reflex of Classical Greek /ĕ/; (2) the obligatory occurrence of /n/ after otherwise vowel final words; (3) the replacement of /f/ by /p/ after /s/; (4) the occurrence of [š] as the alveo-palatal allophone of /x/ before /i/ and /e/ in contrast to palatal [x̭] of standard modern Greek.

The Cappadocian dialect was spoken up until the end of the last century in western Asia Minor in a small area between Nenist and Kanahisan. Its phonological characteristics were: (1) the loss of unstressed /i/ finally; (2) the assimilation of /e/ to /o/ preceding a syllable in which /o/ occurs; (3) the occurrence of additional vowel phonemes of /ö ü ə/. The last two features are evidence of the great amount of Turkish influence in the phonology as well as in the lexicon.

Southern Italian Greek is the modern Greek dialect spoken in the two southern peninsulas of Italy, Calabre and Apulie. In 1958 Stam. C. Caratzas reported (L'Origin des Dialectes Néo-Grecs de L'Italie Méridionale) that there were over 30,000 Greek speakers in these two regions. One of the communities consists of a cluster of four Greek villages located at the southern tip of Calabre (the westernmost peninsula) in the mountainous country of Aspromonte. The other community (nine Greek villages) is located at the tip of Apulie (the easternmost peninsula) in the territory of Salente. Although the Greek speakers of southern Italy are also bilingual in Italian and have many borrowings in their dialect, their dialect is still mutually intelligible with other modern Greek dialects.

The characteristic phonological features of the Greek dialect of southern Italy are: (1) the loss of /s/ in word final position; (2) the replacement of [l] by [ḍḍ] ; (3) the occurrence of [g] before /a, o, u/ where standard modern Greek has [γ]; (4) the replacement of the sequence /nθ/ by /tt/.

Southern Italian Greek dialects distinguish the Calabre subdialect from that of the Salente subdialect. The characteristics of the Calabre dialect (Bovien) are primarily differences in interphonemic specification from standard modern Greek: (1) the occurrence of /u/ where modern Greek has /o/ as the reflex of Classical Greek /ō/; (2) the occurrence of the sequence /st/ where standard modern Greek has /kt, xθ, or pt/; (3) the occurrence of /ts/ where standard modern Greek has /ps/ and /ks/. The Salente (Terre d'Oranto, Grec Orantin, Gríco) dialect has the following characteristics: (1) the loss of /θ/ which is replaced by a /t/ initially and a /t/ or an /s/ medially; (2) the retention of /d/ where modern Greek has /ð/ as a reflex from Classical Greek /d/; (3) loss of /t, d/ intervocalically;

(4) the occurrence of /kt/ where standard spoken Greek has /p^t/ and /kθ/; (5) the replacement of /ps/ and /ks/ by /fs/.

TSAKONIAN, A SECOND MODERN GREEK LANGUAGE

Tsakonian is spoken in the Arcadian Department on the eastern coast of the Peloponnese of the Kingdom of Greece. Geographically it is bounded by the Parnon (Molevo) Mountains to the west, the Gulf of Argolis to the east, and the Lenidi and Saint Andrew Rivers on the north and south respectively. Although the inhabitants of two towns of this area (Karakovouni and Saint Basile) speak modern Greek, the remainder of the population, concentrated principally in nine towns, consists of over 10,000 speakers of Tsaconia (cp. Pernot., 1934). The five towns of Lenidi, Pramatefti, Melana, Sapounakeika, and Tyros are located in the southeastern section of the region close to the coast. The remaining four towns of Prastos, Kastanitsa, Sitena, and Saint Andrew are located in the northwest part of the region.

The unusual nature of the Greek of this area was first noted in the 15th century by Marjare; who spoke of the 'barbarous speaking' people of Laconia, who are 'now called Tsakonian'. Later in the 16th century Gerlach, the Ambassador of Maximilian II to Constantinople, spoke of the Tsaconians as having the only Greek which was not intelligible to other Greeks. Descriptive analyses of the language have been carried out by Hatzidakis, Anagnostro Paulos, Deffner, Deville, and others. The general concensus of the results of their works is illustrated by Gustave Deville, who undertook to answer the question of the relation of Tsakonian to Modern Greek (in

Etude du dialecte Tzaconien, 1866) and concludes that: "Tsakonian is the heir of that Laconian dialect which was formerly spoken precisely in the same locality [translated]." E. Bourguet (Le Dialecte laconien, Paris, 1927) cites as 'incontestible' the following evidence indicating Tsakonian is the modern development from the old classical dialect of Laconian: (1) the retention of /w/; (2) the reflex /dd/ which corresponds to /zd/ or /dz/ of other dialects; (3) the merger of /θ/ with /s/; (4) the loss of word initial /s/ before consonants. Tsakonian also has features which are not either Laconian or Koine: (1) the reflex of Classical Greek /u/ is /u/ rather than /i/ as in modern Greek; (2) merger of /e/ with /i/ before vowels; (3) the loss of intervocalic /s/.

In previous discussions of the linguistic status of Tsakonian, the historical viewpoint has been adherred to, wherein Tsakonian is considered as the only modern Greek 'dialect' which is derived from one of the ancient dialects (viz. Laconian) rather than from the Hellenistic Koine. However, from the viewpoint of synchronic relationships, Tsakonian's status is seen to be that of a separate language because of the language barrier existing between Tsakonian and Modern Greek. This barrier's existence is confirmed by Hubert Pernot in his extensive study of Tsakonian in 1934: "Tsakonian gives to the Greek who hears it for the first time the impression of a foreign tongue, ... even the simplest phrases are incomprehensible for him; he is like a Frenchman hearing a Breton [translated]." On this basis Tsakonian must be considered a separate language from Modern Greek, even though in a historical sense the two may have formerly been 'dialects' of one language; as the dialects split and descended from one parent language, they became two separate but closely related sister languages.

The Tsakonian language has two dialects, which vary little from each other:

Kastanitas-Sitena, spoken in the towns of Kastanitas, the westernmost town on the Saint Andrew River, and Sitena, a few miles north of Kastanitas on the northwestern tributary of the Saint Andrew River;

Lenidi-Prastos, spoken in Prastos and Karakovonve of the northern towns and in the four towns of the south.

ALBANIAN

8. Albanian (Arber, Arbresh, Arvantis, Arnaut, Škip, Shqip) is the official language of the country of Albania, which in ancient times consisted of Illyrica and a part of Epirus. Situated on the Adriatic coast of the Balkan Peninsula, Albania is bounded to the south by the Pindus Mountains, to the east by the highlands of Yugoslavian Macedonia, and by the Dinaric Mountains on the north; it shares political boundaries with Greece on the south and Yugoslavia on the east and north. The territory in which Albania is spoken extends considerably beyond Albania itself—namely into the more or less neighboring nations (Italy, Yugoslavia, Greece, Bulgaria, the Ukraine, Turkey, and Bessarabia). Albanian speakers who have immigrated to America are concentrated especially in Boston and New York.

About half of the two or three million speakers of Albanian live in Albania itself. In Serbia (Kosmet) Yugoslavia there are a half million speakers of Albanian; in Macedonian Yugoslavia, there are another 100,000. In Greece there are 200,000 Albanian speakers located in Attica (especially around Athens, in Eleusia, and in Menich), in Boeotia, in the Peloponnese, and on the Islands of Euboeia and Andros. The Albanian speakers of Italy exceed a half million; they are located in Abruzzi, Molise, Basilicata, Apulia, Calabre, and on Sicily. Unknown numbers of Albanian speakers live in Turkey, Bulgaria, and Bessarabia, and America.

Albanian vocabulary includes a great many Latin loanwords; but, surprisingly enough, Albanian shows almost no influence from classical

Greek. The problem of the subrelationship of Albanian in Indo-European is obscured by the lack of early records—the earliest records of Albanian date only from the fifteenth century.

Albanian has two principal dialects whose less extreme forms are mutually intelligible (Eric Hamp, 1964). The geographic boundary which roughly correlates with the dialectal boundaries is the Shkumbi River which divides Albania approximately in half. The dialects are:

Geg
Tosk.

The Geg (Gheg, Guegue) dialect is spoken north of the Shkumbi River. It has marked subvarieties especially in the city of Shkoder (Scutari) and neighboring mountains, in Kosovo-Metohiya, and in Borgo Erizza, an isolated village on the coast of Yugoslavia near Zador.

The Tosk (Tosque) dialect—the dialect that serves for Standard Albanian—is spoken south of the Shkumbi River as well as in all the Italian and Greek enclaves. The Albanian spoken in Italy and Greece is most like the subdialect of Tosk that is spoken in the extreme south of Albania (Cameriya); so also, the Albanian spoken in Bulgaria, Turkey, the Ukraine, and Bessarabia is like the south Albanian dialect.

In phonology the two dialects differ somewhat in phonemic inventories: where Geg has velar fricatives /x/ and /ɣ/, Tosk has the palatal stops /k^y/ and /g^y/ (although the /k^y/ has an allophone of [x̯]). Geg has also a contrast between /r/ flap and /r̃/ trill, whereas the Tosk /r/ has both flap and trill allophones for the one phoneme. There is even greater difference in the vowel system, for besides having an additional front vowel, /ɛ/, the Geg dialect also has the contrastive additive components of length and nasalization for two vowel subsystems (see below). The phonological systems of Tosk and Geg as described by Leonard Newmark (Structural Grammar of Albanian, RCPAFL 4, 1957) and Dilaver Berberi (Albanian Phonology, M. A. Thesis, I. U. 1963), respectively, are presented below in tabular form:

Tosk Dialect

p		t		c	č	k^y	k		i	ü		u
b		d		ʒ	ǯ	g^y	g		e		ə	o
f	θ			s	š			h			a	
v	ð			z	ž				plus stress			
m		n			n^y							
		r										
		l	L									
					y							

/L/ is 'alveolar tongue-contracted "dark" lateral'.

Geg Dialect

p		t		c	č	k	
b		d		ʒ	ǯ	g	
f	θ			s	š	x	h
v	ð			z	ž	γ	
m		n			n^y		
		l	L				
		r	r̃				
					y		

i	ü		u
e		ə	o
ɛ			a
ĩ	ü̃		ũ
ẽ			ã

All vowels except /ə / occur with contrastive length.

Albanian was not fully recognized as an Indo-European language until the last decades of the last century, because of (1) the abundance of Albanian loanwords from Latin, Turkish, Modern Greek, and Bulgarian; (2) because its 'old endings . . . were so violently changed' (H. Pederson, 1931). Rask as well as Bopp expressed this view. Franz Bopp by 1854, and August Schleicher (Die Sprachen Europas, 1850) both continued the investigation of Albanian and concluded finally that it was indeed Indo-European. The status of Albanian in Indo-European is today considered to be that of an independent branch (having the defining feature of the Satem group), standing in a linking relationship between Armenian and Balto-Slavic. The details and main sound correspondence of Albanian to the other Indo-European languages were worked out by the extensive investigations of Gustav Meyer between 1880 and 1890. The etymological relationships were later refined by the studies of H. Pederson and N. Jokl.

ARMENIAN

9. In general, the Armenian speech area has its traditional center in the eastern half of Modern Turkey—the area of Asia Minor east of a line formed by a northward extension of the western coastline of Syria. At present there are also colonies of Armenian speakers in the western Turkish regions around Smyrna, Istanbul, Ismid, and Rodosta on the north coast of the Marmara Sea. There are additional groups of Armenians in the nearby Middle-Eastern countries of Iraq, Iran, Syria, Lebanon, and Egypt; and Armenian groups are found in India as well as in Europe and the Americas. However, the majority of Armenian speakers are today concentrated in the Armenian S. S. R. and the adjacent Georgian and the Azerbaijan S. S. R. 's .

More than three and a half million Armenian speakers were reported in 1956; 60 percent of the total (approximately 2. 2 million) live in the U. S. S. R. and around 1. 5 million or 60 percent of those in the U. S. S. R. account for more than three-fourths of the population of Armenian S. S. R. The remaining small fractions of people in the Armenian S. S. R. speak Azerbaijani, Ukranian, Russian, Kurdish, and other languages. A half million Armenian speakers live in each of the adjacent S. S. R. 's of Georgia, and Azerbaijan; the remaining Armenian speakers are clustered in the cities of southern Russia. Outside of U. S. S. R. the 1. 4 million (1956) or 40 percent of the speakers of Armenian are widely distributed in:

Turkey, with over 60, 000 speakers;

Syria, with over 120, 000 speakers;

Egypt, with 100,000 speakers;

Lebanon, with 100,000 speakers;

Iran (especially Teheran), where the number of Armenian speakers is unknown;

United States, with over 175,000 speakers;

France, with 70,000 speakers.

According to Herodatus and Endoxus, the Phrygians were the earlier Indo-European speakers, known later as Armenians, who came into Asia Minor from the area north of Thessaly and Thrace about the time of the fall of the Hittite Empire. If this identification is correct, then the first evidence of the Armenian occupation is the conquest of the Till Garimma Province of the Assyrian Empire by a certain Gurdi (who is identified with the Phrygian Gordios).

According to Meillet and Cohen (Les Langues du Monde, 1952) no decisive evidence has been found to establish a Phrygian-Armenian identity. According to Hans Jensen (Altarmenische Grammatik, 1959), comparative linguistic evidence has been found to establish 'a certain close linguistic connection' between the two.

Subsequent to the fall of the Phrygian Empire the country of the Armenians was subjected at various intervals to the Iranian kingdoms of the Medes (ca. 600 B.C.), the Persians (ca. 500 B.C.) under Darius whose inscriptions first refer explicitly to the 'Arminiya', and later the Parthians who contended with Rome for the area. Subjected also to the conquest of the

Macedonians, the Celtic Galatians (from the Middle Danube), the Arabs, and the Seljuk Turks, no one can doubt that the Armenian experience in in being overwhelmed politically began early.

In modern Armenian the two fundamental groups of dialects are the West Armenian and the East Armenian. Adjarian (Armenische Dialectologie, Emische Ethnographic Sammlungen Band VIII, Moskau, 1911) also sets up a third division of the dialects which he calls the y-l group. A. Abeghian (1936), however, contends that this third group should be classified with East Armenian. The dialects are listed following Abeghian, with the three members of the y-l group being listed at the end of the East Armenian group.

Eastern Armenian, including:

Eriwan

Tiflis

Karabagh

Schamachi

Astrachan

Dschugha (Dschulfa)

Agulis

Choi-Salmst

Urmia-Maragha (Northern Persia)

dialect of the Armenian from Artwin

Western Armenian including:

Karin (Erzerum)

Musch

Van (Wan)

Tigranakert (Diarbekir)

Charberd (Erzenka)

Schabin-Karahissar

Trapezunt

Hamschen (on the Black Sea)

Malatia

Kilikien

Syrien

Arabkir

Akn

Sebaste

Ewdokia (Tokat)

Smyrna

N. Komedia

Konstantinopel

Rodosto

Krim

the dialects of earlier Austria-Hungary Region (Poland, Bukowina Transylvania, and Hungary)

Although the fact of Armenian being an Indo-European language was recognized early—for example, Johan Joachin Schröder (Thesauris Linguae

Armenicae, 1711) points out the 'Aryan' connection of Armenian—the subrelationship of this language to other languages in the family was difficult to establish. Rasmus Rask in 1814, methodically pointing out the subrelationships among all Indo-European languages, assigned Armenian to an independent branch. However, he later bracketed Armenian with Iranian in the same branch; still later, he returned it to the status of a separate branch. In 1837 Heinrich Peterman applied the 'comparative method' to Armenian and thereby confirmed Rask. Franz Bopp (1833) and later Schleicher (1850), in addition to others (F. Müller, Windischman, Lagarde), advocated the earlier position temporarily held by Rask—that Armenian belonged in the Iranian branch. It was not until the 1870's that Heinrich Hübschman finally established the status of Armenian as an independent branch by means of clearly differentiating the loanwords (especially numerous from Iranian languages) from the inherited words descended directly from the parent language.

The morpho-syntactic divergences of Armenian from Indo-European turn out to reflect shared structural or typological features with Turkish languages. Phonologically, Armenian shares the innovation which defines the Satam languages and is, because of this, said by Abeghian (Neuarmenische Grammatik, 1936) to be linked to Balto-Slavic by Albanian. In addition Johan Schmidt and later Giuliano Bonfante point out several important shared phonological innovations of Armenian with Greek. The phonological innovation manifested in the numerals for two and for three present especially diverse developments from the Indo-European proto-forms: parental root *dvo occurs as /jerk?u/ in old Armenian, which is analyzed with /k?/ as a reflex of /dv/ and the /jer-/ as an analogical formation from the numeral for three, /jerek^h/. Another remarkable sound change of Armenian is that its /h/ is a reflex of */p/, as in /hayr/ father.

The phonology of Old Armenian is given by Hans Jensen (Altarmenische Grammatik, 1959). There is some problem as to the nature of one phoneme, /γ/, because it was transcribed as /l/ in Armenian loanwords into Persian, Greek, and Syrian—suggesting perhaps that it was at an earlier period a velar (or dark) l. The vowel type is 3(FB) over N—four contrastive tongue heights, with the higher three each making a front-back contrast. A schwa is added to this vowel system by Jensen, but seems to be predictable by his rules, as it occurs obligatorily in certain consonant clusters and as a manifestation of junctures in certain environments. The phonology of Old Armenian, in tabular form, is as follows:

p	t	c		č	k		i		u	
b	d	ʒ		ǯ	g		e		o	(plus stress)
p^h	t^h	c^h		$č^h$	k^h		ɛ		ɔ	
		s		š	x	h		a		
v		z		ž	γ					
		r	ř							
		l								
m		n								
w				y						

There are two standard Armenian dialects or dialect groups today: East Armenian and West Armenian. The speakers of the western group are scattered throughout Turkey and other of the Near and Middle Eastern countries, as well as Europe and the United States. The majority of the Eastern Armenian speakers are concentrated in Armenian S. S. R., other surrounding U. S. S. R. territories, Northeastern Iran, India, and in the U. S. A. The East Armenian dialect for which the Ararat dialect is the basis, is the official language of Armenian S. S. R.

The most important isogloss which characterizes West Armenian is the occurrence of only two series of stops (voiced and voiceless) each preserving the same five linear distinctions shown above for Old Armenian.

East Armenian retains all the basic contrasts in the stop system of Old Armenian — with stops making the same five linear distinctions in three contrastive series: the voiced, the voiceless aspirate, and the voiceless glottalized.

The difference between the two standard dialects is increased by the fact that the voiceless stops of Western Armenian correspond to the voiced stops and voiceless aspirate stops of Eastern Armenian; the Western Armenian voiced stops correspond to the voiceless glottalized stops of the eastern dialect.

Both dialects retain the nasals and liquids of Old Armenian; both modern dialects have lost the /w/ of Old Armenian. The two dialects are alike also in having four linear distinctions for fricatives /f s š x/, in two contrastive series (voiced and voiceless) in addition to an unmatched /h/. Modern Armenian merely adds an /f/ to the fricatives of the Old Armenian.

The vowel system is also preserved in both dialects, with some merger: the close-open vowels (e/ɛ and o/ɔ) have fallen together producing a five vowel system / i e a o u /. Moreover, there is a phonemic schwa in the modern dialects (unstressed in West Armenian). The phonological systems of the two modern dialects are given in tabular form, after Gordon H. Fairbanks and Earl W. Stevick (Spoken East Armenian, 1958; and Spoken West Armenian, 1958):

West Armenian

p	t	c	č	k	
b	d	ʒ	ǯ	g	
f		s	š	x	h
v		z	ž	γ	
m	n		y		
	l				
	r				
	ř				

i		u
e	ə	o
	a	

(plus stress, on all but /ə/)

East Armenian

p	t	c	č	k	
p'	t'	c'	č'	k'	
b	d	ʒ	ǯ	g	
f		s	š	x	h
v		z	ž	γ	
m		n	y		
		l			
		r			
		ř			

i		u
e	ə	o
	a	

(plus stress)

THE IRANIAN SUB-BRANCH OF THE INDO-IRANIAN BRANCH OF INDO-EUROPEAN

10. Iranian languages are spoken in Iran, of course, but not exclusively so; monolingual Turkish speakers make travel difficult in some parts of Iran without a knowledge of Turkish. It has been estimated that Turkish may be the preponderant language of Iran generally, though we lack solid information to support this estimate. Conversely, Iranian languages are also spoken in Turkey, but as minority languages.

More than a score of Iranian languages are spoken by some fifty million people in a dozen different more or less adjacent countries or states or places:

Iran

Pakistan

Afghanistan

Iraq

Syria

Turkey

Hungary

Persian Gulf Islands

Soviet states (Azerbaijani, Uzbek SSR, Turkmen SSR, Tadzhik SSR, The Caucasus)

Numerically prominent Iranian languages include the following (with numbers indicating speakers by the million):

Persian (20)

Kurdish (5 to 10)

Pashto (12.5)

Balochi (2).

The Iranian languages spoken in Iran are (1) Persian, (2) Luri, (3) Bakhtiari, (4) Mazanderani, (5) Gilaki (Gelaki), (6) Gabri, (7) Kumzai, (8) Gurani. Of these Persian is spoken in two other countries beside Iran. Languages (2) to (8), inclusive, were regarded as dialects in the older literature.

Persian is known generally in modern literature by three regional rubrics:

Tehrani Persian

Afghan Persian (not to be confused with a related but different language, Pashto, spoken in Afghan and sometimes called the Afghan language)

Tajik (Tadzhik).

It is true that these regional rubrics are sometimes used as though they served to distinguish three regional dialects, even though the actual basis of this traditional tripartate division reflects, geographic and political boundaries of Iran, Afghanistan, and Tadzhik SSR, respectively, rather than linguistic boundaries. As Jacqueline Wei points out in a privately distributed paper (Dialectical Differences between Three Standard Varieties of Persian, 1962), Afghan spoken in western Afghanistan is similar to some Iranian dialects spoken in eastern Iran, while Afghan spoken in northern Afghanistan is similar to Tajik dialects spoken in south Tadzhikstan.

Nevertheless, each of the three adjacent countries in which modern Persian is spoken recognizes a Standard 'language', yielding three Standard 'languages'—really three dialects out of many more dialects of Modern Persian, but the three which have the greatest currency in Iran, Afghanistan, and Tadzhik SSR. Each of the three is based on Classical Persian as preserved in poetry written a millenium ago. The Standard 'language' of each country is used in published writing, as newspapers and novels, as well as in letter writing of the more formal kind, as in official communications; it is heard in formal utterances on the radio and in speeches, and is mixed with the colloquial in formal conversations—but not in ordinary conversation. Paralleling the three Standard 'languages' are three colloquial dialects named after towns in the three countries concerned—Tehrani, Kabuli, and Varzobi.

In addition to the eight Iranian languages spoken primarily in Iran, the remaining Iranian languages are spoken primarily outside of Iran: (9) Balochi; two Kurdish languages— (10) Zaza and (11) Kermanji; (12) Tat, (13) Talysh, (14) Ossetic, (15) Yaghnobi, (16) Munji, (17) Sanglechi-Ishkashmi, (18) Wakhi, (19) Shughni, (20) Parachi, (21) Ormuri, (22) Pashto. Some of these languages—(10), (11), (12), (13), especially—were regarded as dialects in the older literature.

Iranian languages are conventionally divided into a Western group and an Eastern group, a division which does not reflect the present geographical distribution of these languages. The Western group includes such languages as Persian, Kurdish, and Luri; the Eastern group includes such languages as Pashto, Ossetic, and the Pamir languages. A central group consisting of Parachi, Ormuri, and Balochi is sometimes recognized. However, Balochi is usually considered Western Iranian and some scholars argue for the eastern affinities of Parachi and Ormuri.

The dozen or more languages of the Western Iranian group are spoken in Iran, Afghanistan, Tadzhik SSR, Azerbaijan SSR, Turkey, Syria, Iraq, Western Pakistan, and the islands of the Persian Gulf. The most conspicuous language of Western Iranian is Persian. Western Iranian may be further divided into a Northwestern subgroup (including Kurdish, Tat and Talysh) and a Southern subgroup (including Persian, Luri and Bakhtiari).

The nine extant languages of the Eastern Iranian group are spoken in Pakistan, Afghanistan, parts of Persia and in enclaves in the central Caucasus and Hungary. The most conspicuous language of this group is Pashto, with about 12.5 million speakers primarily in Afghanistan and Pakistan. The greatest number of extant Eastern languages are members of the Pamir subgroup in the northeastern Afghanistan area. Ossetic is geographically isolated from the other Eastern Iranian languages, and is one of two surviving members of a Northern subgroup, along with Yagnobi.

The languages of the Northern Iranian subgroup (also called Saka-Scythian or Scythian) of the Eastern Iranian group, were spoken by war-like nomads—the first of whom were the Scythians—who began to emerge out of central Asia about 800 B.C. and spread over South Russia as far west as Hungary. The names of these languages and dialects are known from Greek inscriptions in South Russia (except for Khotanese which is known from a Buddhist Manuscript in Chinese Turkestan and Khwarizmian which is known from a few glosses preserved in Arabic books excavated by Soviet archaeologists). Sogdian was discovered in the first part of this century, and Khotanese studies have only begun in the last decade. There are two extant representatives of this subgroup: Ossetic, which was probably part of the Alanic complex; and Yagnobi, which is a probable descendent of Sogdian. Among the more important extinct languages of this subgroup are:

*Scythian
*Sarmatian
*Alanic
*Sogdian, the pre-Islamic lingua franca of Central Asia,
*Khwarizmian, which seems to be related to Avestan,
*Khotanese (Khotanese Saka).

Hermatta (Iranian Tribes in South Russian, Budapest, 1952) has postulated extensive dialect variation in the Northern subgroup, beginning in the 6th century B. C.; a number of separate languages and dialects are supposed to have existed in the first centuries A. D.

The Iranian languages are now listed.

(1) Persian is spoken by about 20 million people, primarily in Iran, Afghanistan, Tadzhik SSR, Iraq, and the islands of the Persian Gulf. The most concentrated area of distribution is in a belt from the western portion of Iran eastward through Afghanistan, and northeastward into Tadzhik SSR.

Throughout this area there are many regional dialects, but literary Persian is read everywhere by educated speakers, and the three regional standards, Tehrani Persian (in Iran), Afghan Persian (in Afghanistan), and Tajik, are based on Classical Persian.

Tehrani Persian has been heavily influenced by the Khorasani (Dari) Persian dialect of Afghanistan, as it was from this area that Persian liberators came to drive out the Moslem conquerors who had occupied the greater part of Iran. The present distribution of Tehrani Persian and its local varieties, is in central and southcentral Iran. The following list indicates the prominent local variations:

Qazvini

Mahallati

Hamadani

Kashani

Isfahani

Seden

Kermani

Araki

Shirazi

Jahromi

Lari

Tangistani

Shahrudi
Kazirauni

Mashadi (Meshed)

Tehrani Persian is familiar throughout Iran as the language of education, government, and literature: educated speakers of the regional dialects and languages of Iran are generally fluent in Tehrani Persian. Another variety, but not merely a local variation, is Judeo-Persian (written in Hebrew script) which is spoken throughout Iran, and heavily colored by the local dialects. The Hebrew loans in certain varieties of Judeo-Persian preclude immediate mutual intelligibility with Tehrani Persian.

There are also a number of dialects which are not readily intelligible to a speaker of Tehrani Persian, though such dialects may form chains of intelligibility:

Na'ini

Anaraki

Shushtar-Dizfali

Bihbihani

Bandar-Abbasi

Yazdi
Birjandi

Semnani

Biabanaki

Zarandi

The Standard Persian of Afghanistan differs considerably from that of Iran, on the one hand, and from that of Tadzhik SSR, on the other. Afghan Persian is spread throughout the northcentral portion of Afghanistan from the

Iran border on the east to the Pamir Mountains in the northeast on the border of Tadzhik SSR to the north and Pakistan and Kashmir across the mountains to the south. Persian and Pashto are the two national languages of Afghanistan, even though Persian speakers outnumber Pashto speakers. Among non-Persian speakers bilingualism is very high, particularily for speakers of Uzbek and Persian (or Tajik) and of Pashto and Persian (or Tajik). The Afghanistan dialect of Persian with the widest distribution is Khorasani (Dari) in the provinces of Herat, Ghor, Ghazini, and Hazarajat. There are several regional or tribal varieties of Afghan, among which are the following:

Firozkohi
Djamchidi (Yemchidi)
Timuri (Taimouri)
Taimani

Hazaras, in the south Hindu Kush (including the Berberi, or West Hazaras), speak Persian—some with a liberal mixture of Mongolian and Turkic loans.

The Standard Persian of Afghanistan has both a formal and an informal style. The formal style is modeled on Literary Persian, and may be closer to Tehrani Persian, in general. The informal is based on the local Persian colloquial, and may be closer to Tajiki, in general.

Tajik is spoken in the Tadzhik SSR and in northeastern Afghanistan (in Budakhshan, Panjsher, and Kabul). The Tajik dialects of the various mountain valleys show considerable dialect differentiation. The Tajik which is spoken on the plains is relatively uniform: it differs considerably from the mountain valley dialects. A very divergent form of Tadzik, called Galcha, is found in the Pamir Mountains.

Standard Tajik varies considerably from both Farsi and Afghan Persian; whereas Farsi and Afghan Persian are readily intelligible, Farsi and Tajik, and to a lesser degree Afghan and Tajik, are intelligible only with difficulty. Hodge (Spoken Persian, Washington, 1960) says that there is a considerable difference between the Persian spoken in Tehran, and that in the Tadzhik SSR; still, "this is not a question of boundary lines but of increasing differences in the language as one goes from one place to another within the area where the language is spoken. For example, Afghan Persian is quite close to Tajik, Meshed Persian is more similar to Afghan Persian, and so on..."

According to Hodge (personal communication), Tajik Persian and Tehrani Persian would be the pair to choose if anyone were inclined to bifurcate all Persian dialects--rather than Afghan Persian and Tehrani Persian. Of the three terms, 'Tajik' is often used ambiguously, as though it were a cover-term for any kind of rustic or backwoods or peasant or non-urban Persian. The term is not so used in this report. Tajik Persian is as clearly distinguishable from Tehrani Persian as is Afghan Persian--and, if anything, more easily distinguishable. Each of these three major dialect groups of Persian includes a variety of subdialects, as indicated above; each is spoken not only by native speakers but also as a second language (especially Tajik Persian)--by bilingual speakers; each is spoken both in Iran and in adjacent countries.

The next group of Iranian languages, (2) to (8), inclusive, are less well known, and less widely distributed; they are confined to Iran.

(2) Luri, a tribal group of southwestern Iran;

(3) Bakhtiari, a tribal group in the mountain country of Southern Persia, which is divided into two tribal groups:

Haftlang

Chaharlang;

(4) Mazanderani, in northern Iran;

(5) Gilaki (Gelaki), in the Gilan region (less than 100,000 people);

(6) Gabri, spoken by Persian Zoroastorians in their personal communications as a private language, in the Yezd and Kerman areas in Iran. Gershevitch also includes the following as independent Iranian languages:

(7) Kumzai,

(8) Gurani.

The remaining Iranian languages are spoken outside of Iran.

(9) Balochi (Baluchi), is a language spoken by about two million people in the southwestern part of West Pakistan (1,000,000); in southern Afghanistan (200,000); in the eastern border regions of Iran west of Pakistan and Afghanistan (600,000); in an enclave in the Turkmen SSR in the Merv region (8,000); in areas of north Pakistan; in India (50,000); and in the islands of the Persian Gulf and the northern coast of the Arabian Peninsula (15,000). Balochi in West Pakistan constitutes the predominant language of Kalat and

Quetta; it is also prominent in Western Baluchistan, Khurasan, and Sistan in Iran.

Balochi closely resembles modern Persian in structure (with many borrowings from Persian and Arabic, and in the north from Turkoman), but is an independent language (not a dialect of Persian).

There are two Balochi dialect groups:
Eastern, east of Quetta;
Western, including Makrani.

In West Pakistan, the two Balochi dialects appeared to be separated by a wedge of Brahui, a Dravidian language separated from the rest of the Dravidian family by more than eight hundred miles of Indo-Iranian territory. But as Murray B. Emeneau has recently shown (P-APS 106. 430-42, 1962), it is not realistic to speak of a Brahui wedge between two dialects of an Iranian language (Balochi); rather, the realistic view is that of interpenetration between a language belonging to one branch of the Indo-European family (Balochi) and one branch of another family (Dravidian) which is wholly unrelated to Indo-European. This is not an isolated instance of such linguistic interpenetration in South Asia; rather, it is a special example of a widespread phenomenon of bilingualism, which is counted as having low value in Hindu culture because high value is usurped by Sanskrit ('language of the gods'). Nevertheless, although not viewed with awed respect, bilingual situations are commonly encountered in South Asia, and most interestingly so when one of the two languages shared by a given speaker belongs to one language family, the other to another language family. Emeneau argues that "...evident Dravidianization of Sanskrit in some of its structural features must lead to the partial conclusion that...generations of Sanskrit speakers learned their Sanskrit from persons whose original Dravidian linguistic traits...provided the model for succeeding generations." (p.434). From this, Emeneau turns to the modern opportunity for Dravidianization of Balochi, and vice versa--namely bilateral bilingualism. It is possible that there may be a higher proportion of mother-tongue Balochi speakers who are bilingual in Brahui than of mother-tongue Brahui speakers who are bilingual in Balochi. Unreliable though exact percentages may be, the picture of bilateral bilingualism emerges, in contrast to the unilateral bilingualism which seems to characterize the bilingual situation in most parts of the world. The census of 1901 reports that the Khan speaks Brahui officially and Balochi domestically; but it is difficult to say, in an instance of bilateral bilingualism, which of the two languages involved bears the greater prestige. Whatever its history, the linguistic outcome is clear: '...Brahui...has borrowed several structural features from Balochi...'(p.440).

(10) and (11). Kurdish is conventionally divided into two dialect groups separated by a language barrier; but not complete concensus that these represent two separate languages:

Northern (Western) and
Southern (Eastern),

with the boundary roughly coinciding with a line drawn from Lake Urmia to the junction of the Greater Zab and the Tigris Rivers. Elizabeth Bacon

(1964) says that the Northern and Southern groups are mutually unintelligible. Likewise, Ilya Gershevitch lists Zaza (Northern Kurdish) as a separate language. However, L. O. Fossum (A Practical Kurdish Grammar, 1919) cites the Northern Kurdish specialist , Lerch, as saying: "In general the Zaza Kurds also understand the Kermanji [Southern Kurdish]." Following Bacon and Gershevitch we list the two Kurdish groups separately:

(10) Zaza (Northern, or Western Kurdish)

(11) Kermanji (Southern, or Eastern Kurdish)

Sulaimaniya (Southern) is the prestige dialect used by the Central Government in Baghdad for Kurdish textbooks for elementary schools throughout Iraqi Kurdistan, and for Kurdish broadcasts from Radio Baghdad; the United States Information Service uses this dialect in its weekly Kurdish language news bulletin.

The two Kurdish languages total about 5 million people (Kurdish nationalists claim 10,000,000) located in northern Syria (250,000), Iraq (900,000), eastern Turkey (2,000,000), northwestern Iran (1,000,000), and a small section in Soviet Armenia (several thousands). The area they inhabit forms an equilateral triangle with its apex about at Erevan in the Armenian SSR in the north, and in the south its base is in the extreme north of Syria in the east and across the adjacent Iraq extending west into Iran. They occupy the great arc of the Zagros and Taurus Mountains extending from about Kermanshah in western Iran through northeastern Iraq to Erevan in the Armenian SSR to the north and Erzmcan in eastern Turkey to the west. The area in which the Kurds live is often called Kurdistan (Kurdestan, Kordestan), and the Kurds have long fought for the autonomy of this plateau and mountain area bordering the five nations listed above. Although the area is mostly Kurdish, there are minority communities of Persians, Aramaic-speaking Syrian Christians and Turks. There are also Kurds outside of this area, such as around Aleppo in Syria, and the Kurds of Northern Iran transplanted by Shah Abbas the Great in the late 16th century.

The most recent and comprehensive study of Kurdish is by D. N. MacKenzie (Kurdish Dialect Studies-I, II, 1961); MacKenzie divided the Kurdish dialects of Iran and Iraq into the two groups listed below.

Group I dialects (Southern):

Suleimaniye, in the neighborhood of the town of Suleimaniye in Iraq;

Warmawa, in eastern Iraq between the towns of Suleimaniye and Hulebje;

Bingird, in the neighborhood of the town of Bingird on the Little Zab River in Iraq;

Pizdar, along the northern bank of the Little Zab River near Qal's Dize in Iraq and perhaps across the border into Iran;

Mukri, in an area in western Iran south of Lake Urmia;

Arbil, around the town of Arbil (Irbil) in Iraq;

Rewandiz, around the town of Rewandiz (Rawanduz), in northeastern Iraq;

Xosnaw, in northeastern Iraq south of Rewandiz (Rawanduz), east of Arbil (Irbil), and north of Koy Sanjaqu;

Surci on both sides of the Great Zab River between the towns of Akre and Rewandiz (Rawanduz) in northern Iraq.

Group II dialects (Northern):

Akre, around the town of Akre in northern Iraq;

Amadiye, around the town of Amadiye in northern Iraq;

Barwari, in northern Iraq north of Amadiye and perhaps across the border into Turkey;

Gulli, in northern Iraq northeast of Zakho and perhaps across the border into Turkey;

Zakho, around the town of Zakho in northern Iraq;

Sheikhan, around the town of Sheikhan in northern Iraq;

Dohuk, near Sheikhan.

(12) Tat is spoken by less than 100,000 people on both slopes of the eastern extension of the main Caucasian chain and on the Apsheron Peninsula, in the coastal zone of the Aspian Sea between Apsheron and Kuba, and in Azerbaidzhan SSR and Daghestan ASSR. The Tat are divided into Muslim Tat (60%) and Jewish Tat (40%). The Jewish Tat are concentrated in the area of Kuba, Derbert, and Buinakak with other settlements in areas of Groznyi, Nal'chik, and on the Kuban. The speech of the Muslim and Jewish Tat represent two distinct dialects.

(13) Talysh is spoken by an estimated 100,000 to 150,000 people in the USSR and 50,000 people in the Talysh region of Iran, and in Russia it is spoken along the Caspian coast south of the Viliazh-chai River in the Lenkoran', Zuvand, Astara, and Massala rayons of Azerbaidzhan SSR and extends into Persian territory up to Kepri-chal. It has also been suggested that all the

Talysh in the USSR are bilingual in Azerbaijani Turkish.

(14) The Ossetic dialects are located in the valleys of the Central Caucasus and on the adjacent plains to the north and south of the central chain of the Caucasus. To the north Ossetic is located along the Terek River and its affluents to the west, the Gizel'don, Ardon, and Urukh Rivers in the North-Ossetian ASSR; in the South it is located in the adjacent areas of the Georgian SSR in the South-Ossetian AO; smaller groups are also located in the south beyond the Kura River, and one small enclave is located to the southeast of South Ossetia. In 1939 there were about 354 thousand speakers of Ossetic, excluding Jassic (Russian Census). The Jassic dialect is located just east of Budapest in Hungary.

There are three Ossetic dialects:

Iron (Northeastern or Eastern, Ossetic, Ir, Tagaur, Alagir, Kurtat), spoken in the area of the Fiagdon, Gizeldon and Ardon, tributaries of the Terek River, and also along the middle course of the Terek River;

Digoro-Tual (Northwestern or Western Ossetic) with two subdialects: Digor, in the region of the Urux River, and Tual (Southern Ossetic), which is found in the south among the Georgians.

Jassic, located between Budapest and the Tisza River, centering around the city of Jaszbereny, as well as in a few scattered villages around Hungary.

The Northwestern dialect possesses a written literature in Cyrillic script, in which a number of folkloristic texts have been recorded. The Jassic dialect is cited by J. Nemeth (Eine Wörterliste der Jassen..., Berlin, 1959, p.

p.28) as being not too divergent from the Ossetic dialects of the Caucasus. Jassic and Digoro-Tual, in fact, bear a closer relationship to each other, than either does to Iron.

(15) Yaghnobi is spoken in the valley of the Yaghnob, at the headwaters of the Zarafshan River, considerably north of the Pamir languages.

The languages of the Pamir (Ghalchah) sub-group of Eastern Iranian are spoken in extreme northeastern Afghanistan and adjacent areas in the Soviet Union and West Pakistan. The Dardic languages are located immediately to the southwest. The Pamir languages follow below (16 to 17):

(16) Munji (Munjani) is spoken in the Munjan area of Afghanistan (about four dialects) and in the upper Lutkoh Valley of Chitral, Pakistan (one dialect). Yidgha (Yudgha), the dialect spoken in Chitral which is the most widespread Iranian language in the area, is spoken by 200 to 300 families.

(17) Sanglechi-Ishkashmi (two dialects) is spoken in the Sanglech Valley and the Ishkashim area, Afghanistan.

(17b) Zebaki, which may be a dialect of the Sanglechi-Ishkashmi.

(18) Wakhi is spoken in Afghanistan east of Ishkashim, and still further east in the upper Yarkhun Valley of Chitral, Pakistan, by less than 5,000 people.

(19) Shughni (Shighni, Khugni) is spoken on both sides of the Afghan-Soviet border, some 30 miles north of Ishkashim. Sarikoli, spoken east of Shughni in the Soviet Union, may be only dialectally different from Shughni.

(20) Parachi is spoken near Kabul, in Afghanistan: in Hindu Kush Valley

Shutul (400 families), in Ghujulan (100 families), Nijrau and Tagau (600 families), and in Pachaghan. Dialect diversity seems to be slight. Georg Mongenstiernne (Report on a Linguistic Mission to Afghanistan, Instituttet for Sammenlignende Kulturforskning serie 1-2) suggests the close affinity of Parachi and Ormuri (below), and, further, that the two languages are 'essentially' Eastern Iranian.

(21) Ormuri (Bargista) is spoken slightly north of latitude 32°N, longitude 70°E in eastern Afghanistan, surrounded by Pasto-speaking people. There are two dialects:
Kanigurami, in Wazirstan near Kaniguram
Logar, in the Logar Valley.

(22) Pashto (Afghan, Passtoo, Pakhtoo, Pushto, Pashtu, etc.) is spoken by about 12 million people in Pakistan (5,550,000) in the districts of Peshawar, Hazara, Kohat, Bannu, Dera Ismail Khan, in the Territories of Swat, Buner, and Bajaur, and in northeastern Baluchistan (in the Quetta-Pishin, Lora lai, Zhob, and Sibi areas), and in the Punjab area in the Afghan-Pakistan border areas around Mianwali; in the East, South, and Southwest of Afghanistan (6,500,000) by most speakers in the Eastern, Southern, Kandahar and Farah Provinces, by over half the speakers in Kabul, by one-third of the speakers in Herat, Qutaghan, and Mazar, and in enclaves in all other areas except the northwestern provinces of Maimana, the extreme northeastern Province of Badakhashan, and the Daisangi District of the Province of Kabul.

Herbert Penzl (A Grammar of Pashto, Washington, 1955) succinctly summarized the history of Pashtu studies: "It is not suprising that, at first, the lack of material revealing earlier stages kept scholars from recognizing Pashto as an Iranian language. It is due to the comparative work of W. Geiger, J. Darmesteter, and Georg Morgenstierne that the East Iranian character of Pashto was established. The setting up of phonetic laws describing the development from Proto-Indo-Iranian to Pashto, and the etymological derivation of a sizable part of the Pashto vocabulary from word-forms found in Avestan mark the conclusion of this achievement. The frequent and constant influx of loan-words from literary Persian and colloquial Afghan Persian into Pashto, and the great syntactical influence of Persian upon Pashto cannot obscure great structural and certain significant historical differences between the two Iranian languages."

The dialects of Pashto are often divided into an Eastern (Northeastern, Pakhto), and a Western (Southwestern, Pashto), e.g. by the Linguistic Survey of India. Penzl lists three regional Standards listed from west to east:

Kandahari (Pashto)
Eastern Afghan dialects (Pakh'to)
Peshwari (Pakhto).

Among the local dialects of Pashto the following are listed in the various sources:

Mohmandi
Ghilzai

Durani

Yasufzai

Afridi

Kandahari

Waziri

Kati

Chinwari

Mangal

Wenetsi.

IRANIAN PHONOLOGIES

Persian (including Tajik) as described by V.S. Rastorgueva (A Short Sketch of Persian, IJAL 30, 1964, and A Short Sketch of Tajik, IJAL 30, 1964); Carleton T. Hodge (Spoken Persian, Washington, 1960, and Some Aspects of Persian Style, Lg. 33, 355-69, 1957), Lotfollah Yarmohammadi (A Structural Analysis of Modern Persian, MA Thesis, Indiana University, 1962), and Serge Oblensky (Persian Basic Course, Washington D.C., 1963), distinguishes the following phonemes:

p	t	č	k	q			
b	d	ǯ	g				
f	s	š	x		h	i	u
v	z	ž				e	o
m	n					ae	a
	l						
	r						

All investigators but Yarmohammadi indicate /y/, but Yarmohammudi sets up [w] and [y] as non-vocalic allophones of /u/ and /i/.

For Tehrani Persian Hodge and Yarmohammadi include the glottal stop /ʔ/.

For Tajik Rastorgueva includes /c/, though indicating that it is 'marginal'.

/w/ is a phoneme with limited distribution (see Hodge, etc).

Tajik and Afghan Persian and the Kermani dialect have a separate phoneme /ɣ/.

Hodge (Lg. 33) included /ə/, but it is evidently limited in distribution, and occurs only in informal speech; stress /ˊ/, and four degrees of pitch /1234/.

Hodge indicates the following allophony:

/p t/are aspirated before vowels, and released in syllable final position;

/k g/are aspirated before back vowels /a o u/, but palatalized before front vowels /i e ae/ and in syllable final position;

/q/ is a back voiced velar stop [ġ] initially, back voiced velar fricative intervocalically [γ̣]; [ġ] occurs in /Cq/ and gC/ clusters, and medially as /qq/, and in final position; otherwise [γ̥];

/x/, is a back voiceless velar [x̥], which sometimes occurs with a uvular trill;

/h/ occurs as lengthening of the preceding vowel in rapid speech;

/r/ is a two or more flap-r, voiceless initially;

/ʔ/ occurs only in absolute initial position in informal speech, and has an allophone of glottal stricture ['] elsewhere, though in formal speech [ʔ] replaces ['] also;

/m n/, when occurring in post-vocalic and preconsonantal position, mark the nasalization of the preceding vowel, with assimilation to the articulatory position of the following consonant; otherwise /n/ is dental and /m/ labial nasals;

/w/ occurs in the sequence /ow/, /Vwc/ (where V = /eau/);

/i a u/ occur slightly lengthened in prestress position;

/ə/ is a phoneme of limited distribution.

All vowels are short in stressed position unless lengthened by a following /ʔ/ in phrase-final position.

The following inventory for Western and Eastern Baluchi (Balochi) is from Grierson (The Linguistic Survey of India), William Geiger (Lautlehre des Baluči, Munich, 1891), and V.A. Frolova (Beludzhskij Jazyk, Moscow, 1960):

p	t	č	k
b	d	ž̧	g
	s	š	
	z	ž	
m	n		
	l		
	r		
w		y	

i	u	i·	u·
a		e·	o·
		a·	

In the western dialect of the Soviet Union Frolova includes the retroflex stops /ṭ ḍ/ and the fricatives /x γ/ not present in the other western dialects, though in the other dialects (Western and Eastern) a pharyngeal /h/ occurs which is lacking only in the Khorasan and the Soviet Union areas. The retroflex stops /ṭ ḍ/ and their aspirate counter-parts /ṭ' ḍ'/, as well as the aspirates /p' t' č' k'/ occur in the Eastern dialect. (According to Frolova aspirates occur voiced /b' d' z̧' g'/, but only in loanwords.) Fricatives are stop allophones in the Western dialects, but are contrastive in the Eastern dialects. All give /f θ ŏ x γ/, and Geiger adds /v/. Geiger also includes /ṛ/ as occurring in Indic loans, and Frolova adds/w'/ for the Eastern dialects.

Both the phonemic systems presented by McCarus and MacKenzie for Kurdish have the following consonant phonemes:

p	t	č	k	q		
b	d	ž	q			
f	s	s	x		ħ	h
v	z	z				
m	n		n			
	l ɫ					
	r,ř					

To this McCarus adds a glottal stop /ʔ/ and an alveolar /s̲/, the latter being a rare phoneme, as is the phoneme /v/.

MacKenzie describes the glottal stop as predictable, but uses some morphological criteria.

/l/ is a dental, 'light' l, while /ɫ/ is, according to MacKenzie, a 'voiced velarized dental', and, according to McCarus, a 'dark' alveolar.

/r/ is a flapped liquid and /ř/ a trilled liquid.

/ħ/ is a voiceless pharengeal fricative, /h/ a voiceless glottal fricative, and /'/ a voiced pharyngeal fricative.

In the vowels MacKenzie establishes two coexistent vowel systems with contrastive length:

i	u	i·	u·
ø	o	e·	o·
		a·	

where /ø/ is a dipthong [ae] + [ɛ] which is analyzed as a unit phoneme.

McCarus does not establish length as contrastive, but instead postulates the following vowel inventory:

i	ɨ	u
I		U
e	a	o

/i e a/ have both short allophones (in unstressed word final position) and long allophones (elsewhere)*;

/u/ has a long allophone as its only member, and /o/ has a short allophone as its only member.

MacKenzie indicated the phonemic differences of a number of dialects:

Group I dialects:

Warmawa lacks /'/

Bingbird and Pizdar and Mukri dialects lack /ŋ/; and /ć/ and /ź/ (aleolopelatal affricates) occurring for some speakers rather than /č/, /š/. For these speakers [č] and [š] are rather allophones of /k/ and /g/.

In the Arbil and Surci dialects the /ć ź/ alternation is consistent, and the phonemes /ŋ/, /ɨ/, and /ø/ are lacking.

Group II dialects:

Four dialects, Bitlis, Akre, Amadiye and Zakho have the emphatics /i s z/.

Three dialects have contrastive aspiration: /p^h t^h c^h k^h/ in the Erevan dialect, and /p^h t^h k^h/ in the Amadiye and Zakho dialects.

The group two dialects lack /ɨ/.

The Akre, Amadiye, Bitlis, and Zakho dialects have two coexistent phonemic systems:

i		u	i·	ü·	u·
	a		e·		a·

with /ü/ being replaced by i· by some speakers.

In the Erevan dialect there are coexisting vowel systems with an additional contrast in the short vowels, and a 2(FB) over neutral long vowel system:

i	u	i·		u·
a	ö	e·		o·
			a·	

The phonemic inventory of Ossetic is attested by the following sources: W. Miller (Die Sprache der Osseten, Grundriss der Iranischen Philologie, Strassburg, 1903), H. Hübschman (Etymologie und Lautlehre der Ossetischen Sprache, Strassburg, 1887), A. Christensen (Textes Ossetes, Copenhagen, 1921) and Hans Vogt (Le Système des Cas en Ossete, Acta Linguistica 4. 17-41, 1944).

The following consonant inventory is agreed upon by all of the above mentioned sources:

p	t	c	č	k	q
p̓	t̓	c̓	č̓	k̓	
b	d	ʒ	ǯ	g	
f		s	š		x̧
v		z	ž		γ
m		n			
		l			
		r			
w			y		

All sources (but Hockett) add the contrastive glottalized series of stops to the plain aspirated stops, /p t c č k/, but Hockett interprets the plain series as glottalized and adds a constrastive aspirate series.

All sources include the post-velar /q/ in the inventory though Hockett describes it rather as /ʼ̣/, a pharyngealized glottal catch, which 'apparently sometimes has a very far back dorso-velar closure'.

Hockett cites only one source, Hans Vogt, who postulates the inventory as it stands above.

A component of palatalization is combinable with all stops, and fricatives except labials and post-velars.

The vowel inventory is given by Miller as:

i   u   i·   u·
e ə o
æ   a

Christensen indicates a 2(FB) over neutral with an SGC of length.

Hans Vogt, working with the old materials, indicated the following coexistent systems:

i   u   i·   u·
e   o   e·   o·
æ   a   a·

Allophonically, /n/ and /l/ have three variants each: [n], [n^y], [ŋ], and [l], [l^y], [ɫ], respectively; /r/ has two variants: [r:] amd [r^y].

Dialectally, the sibilants /s z/ and the affricates /č ǯ/ vary with the prepalatal /ś _ź/ and /ć ʒ́/; Tual has /ɛ/ rather than /æ/; Iron is the only dialect having /ə/ as a phoneme; and Jassic probably lacks the glottalized stop series, not being in contact with Caucasian speakers. (In one recorded instance an Ossetic /k/ corresponded to a Jassic /k^h/.)

Phonetically, [h] occurs only in interjections.

The phonological systems of the Ghalchah languages are fairly uniform. All the languages appear to have the following consonants (after the Linguistic Survey of India):

p	t	č	k	q
b	d	ǯ	g	
f	s	š	x	h
v	z	ž	γ	
m	n			
	l			
	r			
w		y		

Geiger (Grundriss der Iranischen Philologie) included /c ʒ q/.

For Shighni and Wakhi add / θ δ/. Some retroflex consonants occur in Munji, Wakhi and Zebaki.

All the languages have contrastive length.

Shighni and Zebaki have 2(FB) over N:

i		u
e		o
	a	

plus length

Munji has the coexistent systems of 2(FB) over N (short), and 2(FB) (long):

i·	u·
e·	a·

Wakhi has the coexistent systems of $2(FF^{o}B)$ over N (short), and 2(FB) over N (long):

i	ü		u	i·		u·
e	ö		o	e·		o·
		a			a·	

Sarikoli has the coexistent systems:

i	ü	u	i·		u·
e		o		a·	
	a				

The phonemic inventories for Parachi and Ormuri (taken from Georg Mongenstierne, Indo-Iranian Frontier Languages, Vol. 1, Oslo, 1929) are very similar, and share the following consonant inventory:

p	t	c	ţ	č	d	(q)	
b	d	z	ḍ	ž̧	g		
f		s		š	x		h
v		z		ž			
m		n	ṇ		ŋ		
		l					
		r					
				y			

/q/ occurs only in loanwords.

Parachi also has a glottal stop /ʔ/ and voiced and voiceless aspirates /p‘ t‘ ţ‘ c‘ k‘ b‘ d‘ g‘/; and perhaps / m‘ n‘ r‘ l‘ /.

Parachi and Ormuri both have contrastive vowel length, but whereas Ormuri has a 3(FB) short vowel system and a 2(FB) over N long system, Parachi has a 2(FB) over N (short), with an additional /ə·/ (long).

Ormuri					Parachi					
i	u	i·	u·		i		u	i·		u·
e	o	e·	o·		e		o	e·	ə·	o·
æ	a		a·			a			a·	

THE INDIC SUB-BRANCH OF THE INDO-IRANIAN BRANCH OF INDO-EUROPEAN

11. More than 350 million people in South Asia speak more than a score of different Indic languages. Only five of these Indic languages are not spoken in India proper, or not spoken primarily there:

Lahnda and Sindhi are spoken for the most part in the Indus River Valley of Pakistan;

Singhalese-Vedda is spoken in Ceylon;

Maldivian is spoken on the Maldive and Laccadive Islands off India's southwest coast;

Romany is spoken throughout Europe and in America.

Indic languages are also spoken in South Africa, the South Pacific, the South American Guianas (Western Hindi) and the Antilles, Afghanistan (Lahnda), and on the Andaman and Nicobar Islands. One can distinguish the list of places in which Indic languages are spoken, given in the first paragraph, as places where Indic languages are known to have been spoken from before the 19th century. The other places where Indic languages are 'also spoken', given in this paragraph, are places to which speakers immigrated, characteristically as laborers under the impact of 19th century plantation culture. Such immigrant speakers do not always represent minority languages in their adopted countries or islands, however. For example, speakers of Indic languages in Fiji outnumber the aboriginal Fijian speakers in modern Fiji. It might be useful to speak of overseas Indians, on the analogy of overseas Chinese.

Indic languages are generally grouped into a number of zones, which roughly indicate the internal relationships of Indic. These are listed below, as adapted from George Grierson (Linguistic Survey of India, 1901), Suniti Juman Chatterji (Languages and the Linguistic problem, Oxford Pamphlets on Indian Affairs No. 11, 1945), and J. Raymond Firth (Encyclopedia Britannica, 1964):

Romany—one to three languages: (1) European Romany, Armenian Romany, and Asiatic Romany;
Sinhalese-Maldivian—three languages: (2) Sinhalese, (3) Vedda, (4) Maldivian;
Eastern Zone—three languages: (5) Bihari, (6) Bengali-Assamese, (7) Oriya;
Northwestern Zone—two languages: (8) Lahnda, (9) Sindhi;
Southern Zone—one language: (10) Marathi;
Central Zone—seven languages: (11) Gujerati, (12) Rajasthani, (13) Bhili, (14) Khandesi, (15) Banjiri, (16) Panjabi, (17) Western Hindi;
East-Central Zone—one language: (18) Eastern Hindi;
Northern Zone—four languages: (19) Nepali, (20) Kumauni, (21) Garhwali, (22) Western Pahari.

These zones differ in over-all population. The more populous zones are now listed in descending order, with number of speakers (in the millions):

Eastern Zone (140)
Central Zone (125)
Northwestern Zone
Southern Zone
East-Central Zone
(20 to 30)

Languages in some of the eight zones given above constitute Indic branches or sub-branches, but enough comparative work has not yet been done to justify equating the languages clustered in these eight zones with eight linguistic branches. The zones reflect geographic groupings, of course, but more intricately reflect British colonial policy—or rather policies—which in general did not favor drawing or redrawing boundaries of states in India according to language lines (Marshall Windmiller, Linguistic Regionalism in India, Pacific Affairs 27. 291-318, 1954). After 1949, there was further redrawing and reallocation of state boundaries, sometimes influenced by the distribution of languages in India, by the political difficulty of obtaining a concensus on the adoption of a national language or national languages (partly because of the intermingling of Dravidian speakers and Indic speakers in some states), and by the number of speakers in the most populous languages, as Hindi and Bengali, and the degree of dialect differentiation in a given language which in some cases is extraordinarily high. Beginning with Sir Rabindranath Tagore (1861-1941), for example, Bengali began to be written more or less as it is spoken (in one of its many dialects—that used in Calcutta), besides being written in the literary language formerly used only for poetry—a dialect flooded with Sanskrit words and constructions (Edward C. Dimock, Literary and Colloquial Bengali in Modern Bengali Prose, IUPAL 13 in IJAL 26. 3, 1960). Educated Bengali speakers are agreed in their admiration of English prose, especially as expressed in the English novel, and now produce excellent prose works in Bengali; but they are not yet agreed on whether to present these modern literary efforts in

Literary or Colloquial Bengali, or a mixture of the two. English still serves as a lingua franca in multilingual India, but as formerly it is still restricted to certain classes of Indians, as civil servants. There is considerable political interest in replacing English with one or more Indic languages for this purpose, as Standard Hindi-Urdu; but the Indian elite still speak English. The increasing number of bilinguals for whom Hindi is a second language increases the complexities of the language census in India where, after all is said and done, linguistic states seem not to be in the offing. In lieu of a political basis, and in lieu of any firm modern linguistic basis for grouping Indic languages, we follow the classification by eight zones, which is essentially based on Grierson's information.

At least two other subclassifications have been proposed. E. Benveniste and J. Vendryes segregate non-Dardic Indic languages into five groups (Meillet and Cohen, Les Langues du Monde). A. F. R. Hoernle (A Comparative Grammar of the Gaudian Languages, 1880) postulated what Grierson elaborated - an early cleavage between an Outer Band of languages, and Midland (Sanskrit); after this split, certain Outer Band languages came under the heavy influence of the Midland speech and subsequently became an Intermediate group.

Comparison of these three classifications shows certain similarities, as well as curiously cross-cutting differences. For example, Romany, Sinhalese and Eastern — the first three zones listed above — are also classed as separate groups by Benveniste. Grierson implicitly groups Romany and Sinhalese with his Outer Band languages—the languages of the Eastern, Northwestern and Southern Zones, while Benveniste places the Northwestern and Southern Zone languages and the languages (11) to (15) of the Central Zone in a Western Group. Languages (16) and (17) (Panjabi and Western Hindi) of the Central Zone and the languages of the East-Central and Northern Zones form Benveniste's Central Group. Grierson's Intermediate languages are those of the Central Zone (except (17) Western Hindi), the East-Central Zone and the Northern Zone—i.e. all of Benveniste's Central Group and the Central Zone members of his Western Group. Western Hindi then remains as Grierson's Midland language.

A peculiar difficulty with the zones—partly the reason for treating them as zones rather than as sub-branches—is the lack of strict linguistic boundaries within, and sometimes between, zones. For although areas of homogeneity can be found which are mutually unintelligible with other such areas, there are generally transitional areas of neighboring intelligibility between these focal points. One can find transitional dialects between Western Hindi and Panjabi (in the Central Zone), and between Panjabi and Lahnda (the latter in the Northwestern Zone). Likewise there are transitions from Western Hindi into Rajasthani and from Rajasthani into Gujarati (all in the Central Zone); but also from Rajasthani to Sindhi, the latter just south of Lahnda and like Lahnda in the Northwestern Zone. In the Eastern Zone there is neighbor intelligibility from Bengali-Assamese into Oriya, and perhaps into Bihari. The situation on the Bihari-Eastern Hindi border and the Eastern Hindi-Western border is not too clear, but it is quite possible that the same type of transition occurs. Marathi must be the most isolated of those languages in northern India, for a definite linguistic boundary is reported between Marathi and Gujarati, Rajasthani, Western Hindi, and Eastern Hindi. However, there is the possibility of transitional dialects into Halbi, and thence into Oriya. In fact, Grierson says that one can go from the westernmost parts of Marathi territory through Oriya and into Bengali-Assamese all the way to the easternmost parts of Assam without finding criteria to establish a linguistic boundary. The question here (and throughout the discussion of Indic languages which follows) is whether the difficult-to-find linguistic boundary is tantamount to a lack of language barrier in a given zone of India. In short, the Indic languages surveyed earlier in this century are not mutually intelligible to each other and yet are not entirely attested as separate languages. Despite the existence of language barriers, further field work is needed to determine the linguistic basis of the existing barriers and to discover the nature and function of the transitional dialects.

ROMANY

(1) There are estimated to be around 900,000 Romany (Gypsies) of whom perhaps half speak Romany with some facility.

These Romany are descendants of a group which separated from other Central Indic speaking peoples sometime before the middle of the third century B. C., and remained among the speakers of Northwest Indic and Dardic languages for from eight to eleven centuries before beginning the migration, or migrations, which led them across Persia and Armenia into Europe and eventually into America.

Romany may already have had some dialect differentiation before the speakers of this language left India. Further differentiation occurred after some Romany speakers were left behind in western Asia (the Palestinian Gypsies), and in Armenia; and also as groups separated from each other in their spread over Europe.

In the course of their migrations Romany speakers borrowed vocabulary extensively from the languages with which they had ephemeral contact. It is even possible that borrowing had been so extensive that it rendered the three major 'dialects' of Romany—European, Asiatic, and Armenian—mutually unintelligible; hence there may be three separate Romany languages spoken today. The source of most borrowings in the European subdialects is Greek. But the subdialect spoken as far west as Wales, for example, has in addition loanwords from Persian, Armenian, Rumanian, Bulgarian, Serbo-Croatian, Czech, German and French, besides English.

The major problem in determining the number of speakers of Romany is that of defining the point at which Romany expands itself out of existence. With the retention of a greater or lesser number of Romany (or Romany-borrowed) words, whole groups are known to speak the local language of the country in which the Romany reside. This has led to a distinction being made for two kinds of Romany in some European areas—one kind may be regarded as Romany, and the other as a dialect of the local language with Romany loanwords. For example, this distinction is labelled Romany and Rodi (Rotwelsch) in Norway and Denmark, and 'higher' and 'lower' Romany in Sweden. Among some Gypsies of Wales and the Balkans, Romany remains the native language of the group; but among others, as in parts of England and the United States, Romany has been wholly replaced by the local language (English) with only a few loanwords from Romany.

Another difficulty in determining the number of Romany speakers is sociocultural. Theirs is a fringe society, often feared or rejected by the dominant culture within which they live. Their denigration by dominant peoples has resulted in Romany assuming for the Gypsies the status of a 'secret' language, to be used when gypsies do not wish to be understood by outsiders. The 'secret' would be unguarded if it was taught to non-gypsies; it has been difficult for outsiders to study the language extensively, or even to determine whether it is actually spoken.

SINHALESE-MALDIVIAN

The three languages in the Sinhalese-Maldivian zone are spoken in Ceylon (Sinhalese and Vedda) and on the Islands of Maldive and Laccadive (Maldivian). These languages are separated by clear-cut barriers; there are no transitional dialects. The Vedda and Maldivian are very small groups, the former having less than a thousand and the latter numbering just under 100,000.

(2) Sinhalese (Singhalese, Cingalese) is spoken by 7,000,000 people of whom over 5,250,000 are native speakers. The official language of Ceylon is Sinhalese though Tamil is now also accepted for certain official purposes. The majority of the people in all parts of Ceylon speak Sinhalese except in the districts of Manner, Vavuniya, Jaffna, and Tringomalee in the north, and Batticaloa in the east, and the Nuwara Eliya District in the Central Province.

Sinhalese, with Vedda, is closely related to Maldivian and together these form a relatively well attested sub-branch within Indic. Spoken and written Sinhalese differ sharply from each other, both in grammar and in phonology. For example, in the literary language there are six distinct forms of the present tense as against one in the colloquial. The literary language was once used exclusively for formal writing and preferred for formal speaking, but the colloquial is now used increasingly as a written medium. In its phonology Sinhalese has lost the Indic aspirated stops, but has retained the /y/ lost by some other Indic languages. In morphology Sinhalese conforms to the usual Indic type, but its syntax is said to be influenced by Dravidian. The internal relationships of Sinhalese point most strongly to an earlier connection with the Central Indic language, (e.g. Western Hindi, Rajasthani, Gujarati), as well as to some East Indic influence, (e.g. Bengali, Oriya, Bihari).

The Sinhalese have three 'secret' dialects:

Rodiya (spoken by the Rodiya low-caste group, who live by begging, stealing, and menial labor) has a large number of vocabulary items not found in Standard Sinhalese, including certain words of unknown origin also found in the Vedda language;

Goyi-basava, a dialect used during paddy cultivation to insure the success of the crop by replacing all normal cultivation terms by special terms to 'deceive the demons' and to avoid 'unlucky words';

Kale-basava, used while hunting in the same manner as Goyi-basava.

(3) Vedda (Veddah, Veddha, Veda, Vadda, Vaedda, Wedda, Weddo, Weda,

Bedda, Beda) was spoken by about 803 people in 1953, though less than a decade earlier (1946) Vedda speakers numbered 2, 361. The Vedda language is rapidly being replaced by Sinhalese. The Vedda live in the remote forests east of the central mountains of Ceylon; 370 are located in the Badulla District and 321 in the Polonnaruwa District (one of the two districts which together used to form the Anuradhapura District). Most Indic linguistic sources treat Vedda as a dialect of Sinhalese, but William A. Coates, on the basis of recent fieldwork in Ceylon, reports that the two are mutually unintelligible without 'special training'. The primary differences between Vedda and Sinhalese seem to be phonetic and lexical rather than phonemic, morphemic or syntactic. The Veddas are reported by Geiger, Language of the Vaddas, 1935, as having, in addition to the language proper, special ritual and 'jungle' secret languages. There are some Vedda words for which no Dravidian or Indic etymologies can be found.

(4) Maldivian (Mahl, Divehi Bas) is spoken by nearly 85, 000 people on the Maldive Island chain, and on the Island of Minicoy about 70 miles from the Maldive Islands. Minicoy is part of the Laccadive Islands which are southwest of the Indian subcontinent, as are the Maldive Islands. Maldivian is closely related to Sinhalese but is definitely not a dialect of Sinhalese. William A. Coates points out that they are not mutually intelligible and that there are no transitional dialects. Maldivian has three styles distinguished by a formalized set of honorifics which vary according to the status of the person addressed. Maldivian phonemic distinctions are closely similar to those of Sinhalese. However, Maldivian /f/ corresponds to Sinhalese /p/, Maldivian /š/ to Sinhalese

/ṭ/; and /æ/ and /e/ have merged in Maldivian. There are many Arabic loans in Maldivian.

EASTERN ZONE

The three languages in the Eastern Zone are spoken by nearly 140 million people in northeastern India from the States of Bihar and Orissa to the easternmost part of Assam. Bengali - Assamese has the greatest number of speakers— about 76 million speakers; and Bihari has two-thirds as many.

The Eastern Zone languages developed from the Magadhan Prakrit; they are sometimes subdivided as Western Magadhan (the Bhojpuri dialect of Bihari), Central Magadhan (the Maithili and Magahi dialects of Bihari) and Eastern Magadhan (Bengali-Assamese and Oriya.)

(5) Bihari is spoken by nearly 50 million people in the easternmost portions of Uttar Pradesh in the Gorakhpur and Banaras Divisions, throughout the State of Bihar, and in the Tarai Districts of the lower slopes of the Himalayas in Nepal. About two-thirds of the total number of Bihari speakers are located in Bihar State. Bihari is bounded linguistically by Eastern Hindi in the west, Sino-Tibetan languages and Nepali in the north, Bengali in the east, and Oriya in the south. Within the geographic boundaries of the Bihari area, a number of Dravidian and Munda languages are spoken. The entire Bihari area forms a rough equalateral triangle in northeastern India with its apex at the junction of the States of Madhya Pradesh, Orissa, and Bihar in the south, and with its base in the north running along the Indo-Nepalese border from Sikkim in the northeast to about the midpoint of the Nepal frontier in north central India west of the Uttar Pradesh-Bihar border.

Bihari was recognized as an independent language in the late 18th century but is sometimes erroneously considered as a dialect of Eastern or even Western Hindi. Some writers have treated the Bihari dialects, Bhojpuri and Maithili-Magahi, as separate languages, though it is quite clear that they are mutually intelligible.

There are three rather than two main dialects of Bihari:

Bhojpuri
Maithili
Magahi.

Bhojpuri (Bhojpuriya, Hindusthani, Deswali, Khotla, Piscimas) is spoken by 23,500,000 people in the Gorakhpur and Banaras Divisions of Uttar Pradesh; and in the districts of Champaran, Saran, Shahabad, Palamau, and Ranchi in western Bihar. There are also hundreds of thousands of Bhojpuri speakers in Calcutta. In Bengal there are many Bhojpuri speakers who cannot be accurately enumerated as they refer to themselves as speaking 'Hindustani', and are returned as such in the census. Thus they are confused with the Hindustani speakers of the Western Hindi dialect of that name, i.e. Standard Hindi-Urdu.

Tiwari, in the Origin and Development of Bhojpuri (1960) reports on the language situation at a time when Bhojpuri was still holding its own; it may before long be overshadowed in education and public life by the national language, Hindi-Urdu. Bhojpuri remains the daily language of the home and is even used in the classrooms (for difficult oral examinations and by students in the lower grades in addressing their teachers). Some Pandits in older Sanskrit style schools still use it for instruction. Language loyalty is such that anyone in the villages using a language other than his native Bhojpuri to his own people is looked upon as a pedant or as one given to superior airs. Bhojpuri remains the language of folk politics, social and economic affairs, and even of sermons and speeches. Strong popular support has recently been expressed for the development of literature and drama in Bhojpuri.

The subdialects of Bhojpuri listed by Tiwari are:

Northern Standard Bhojpuri (Gorkhpuri, Sarawaria, Basti), spoken by about 10,000,000 people in the districts of Basti, Gorakhpur and Deoria in Uttar Pradesh, and Champarahin Bihar;

Western Standard Bhojpuri (Purbi, Benarsi), spoken by about 6,000,000 people in the districts of Azamgarh, Ghazipur, Banaras, Mirzapur, and the extreme tip of Faizabad in Uttar Pradesh;

Southern Standard Bhojpuri (Kharwar), spoken by about 7,000,000 people in the districts of Ballia and Ghazipur in Uttar Pradesh, and Saran and Shahabad Districts in Bihar, and considered the Standard for Bhojpuri as a whole;

Nagpuri (Chotar Nagpuri, Sadani, Sadri, Dikku Kaji), spoken by about 2,000,000 people in the districts of Ranchi and Palamau in southwest Bihar.

Tharu, spoken in the Tarai Districts of Nepal;

Madhesi, in the northwestern section of the Northern Standard subdialect;

Domra.

Maithili (Tirahutia), the second main dialect of Bihari, is spoken by about 15,000,000 people in the whole of the districts of Dharbhanga, Musaffarpur, Purnea, Monghyr, and Bhagalpur in Bihar. In the district of Champaran it is current in the eastern part, and merges into Magahi in the eastern part of Patna and the northern part of Santal Paraganas Districts. It is spoken by the people of the Tarai of Nepal on the borders of Bhagalpur and Tirhut Divisions of Bihar, and by the non-Bengali residents of Malda and Dinajpur Districts in Bengal. Subhadra Jha, in the Formation of the Maithili Language (1958) lists seven subdialects of Maithili:

Standard Maithili, spoken in North Darbhanga;

Southern Maithili, spoken in south Durbhanga, east Muzaffarpur, north Bhagalpur, and West Purnea;

Eastern Maithili, spoken in east Purnea, Malda and Dinajpur, with these two areas representing local variations; the latter, spoken in Malda and Dinajpur, is called Khotta;

Chikachiki, spoken in South Bhagalpur, North Santal Parganas, and south Monghyr;

Western Maithili, spoken in west Muzaffarpur, and east Champaran;

Johahi, the language of Moslems in north Darbhanga;

Central Colloquial which has two varieties: that of Sotipura, and that of lower caste people in Madhubani subdivisions.

Magahi, the third main dialect of Bihari, is spoken by nearly 10,000,000 people in south Bihar in the eastern part of the Patna Division, mostly in the north of Chatanagpur Division, and in the Malda District in Bengal. Magahi has three subdialects: Southern, Northern, and Central.

(6) Bengali-Assamese is spoken by more than 76 million people in the extreme northeastern portion of India in the States of West Bengal and Assam in India, and in adjacent East Pakistan.

Bengali (Banga-Bhasa) is spoken by over 70,000,000 people in East Pakistan, in the West Bengal State of India, in the districts of Dhanbad, Manbhum, the northern part of Singhhum, and the south and east portions of Santal Parganas in Bihar State in India; and on the western borders of the districts of Goulpara and Garo Hills in Assam State in India. About 30,000,000 speakers live in India and the remaining speakers live in East Pakistan. Bengali is bordered by Sino-Tibetan languages on the east and north, Bihari on the west, and Oriya on the southwest. Language boundaries are difficult to assign to Bengali because of its close linguistic affinity to the neighboring languages Bihari and Oriya. From village to village it shades off in the southwest into Oriya, and there is an intermediate subdialect area between Bengali and Bihari.

The Bengals have a long literary tradition and an intense language loyalty which has forced Pakistan to acknowledge Bengali as a co-official language with Hindi-Urdu, and has caused considerable resistance to the spread of Hindi-Urdu in the West Bengal State of India.

Grierson lists the following dialects of Bengali:

Central (Standard) Bengali, spoken in West Bengal in the districts of Murshidabad, Nadia, Hooghly, Howrah, Calcutta, 24-Parganas, the northern part of Midnapore, and the eastern half of Burdwan;

Western Bengali, spoken in West Bengal in the districts of Bankura, Birbhum, and the western half of Burdwan, and in Bihar State in Dhanbad, Manbhum, the northern fourth of Singbhum, and the south and east two-thirds of Santhal Parganas, includes the three subdialects:

Kharia Thar, in Manbhum,
Mal Pahoria, in Santhal Parganas and the adjoining portion of Birbhum,
Saraki;

Southwestern Bengali, spoken in the southern two-thirds of the Bengali-speaking area of Midnapore in West Bengal;

Northern Bengali, (spoken in East Pakistan, West Dinajpur in the state of West Bengal, and the northeastern tip of Purnea in the State of Bihar) has two subdialects:

Koch,
Siripuria;

Rajbangsi, spoken in East Pakistan, the districts of Jalpaiguri, Cooch Behar and Darjeeling of West Bengal, and the western borders of the districts of Boalpara and Garo Hills in Assam, includes the subdialect Bahe, spoken in Darjeeling;

Eastern Bengali, spoken in Eastern Pakistan, is divided into two subdialects, and East Central one and Haijong;

Southeastern Bengali, spoken in East Pakistan, includes the subdialect Chakma.

Assamese is spoken by about 6,000,000 people primarily in the Assam State of India, though about 10,000 live in West Bengal. Though Assamese is generally listed as if it were a different language than Bengali, Grierson reports that it is structurally so close to Bengali that 'it would be extremely difficult to oppose any statement to the effect that Assamese was nothing but a dialect of Bengali'. In fact, he notes that the Chittong variety of Bengali is structurally much more different from Calcutta Bengali than is Assamese. The classification of Assamese as a separate language has been perpetuated by a feeling of language loyalty on the part of its speakers; and by the recognition of its distinct literature by outsiders, as well as by Assamese.

Grierson lists the following subdialects of Assamese:

Standard Assamese, spoken in the districts of Sibsagar, Nowgong, Darrang, and Lakhimpur;

Western Assamese, spoken in the districts of Goalpara and Kamrup;

Jharwa, a 'pidgin Assamese' used by Garo tribesmen as a commercial language;

Mayang (spoken in Manipur) could 'with equal (or perhaps more) justice be classed as a form of Eastern Bengali', according to Grierson.

(7) Oriya (Odri, Utkali, Uriya) is spoken as a native language by 13.3 million people in India, of whom approximately 12 million (90%) live in Orissa State. All but about 20,000 of the remaining 1.26 million Oriya live in the four states adjacent to Orissa or a fifth non-adjacent state, Assam: Andhra Pradesh (163,000 Oriya speakers), Madhya Pradesh (304,000), Bihar

(313,000), West Bengal (182,618) and Assam (281,000). The speakers of Oriya who live in the states adjacent to Orissa are primarily concentrated in the districts on the border of Oriya (i.e., Midnapore District of West Bengal); Singbhum and Ranchi Districts of Bihar; Raigarh, Raipur, and Bastar District of Madhya Pradesh; and Visakhapatam District of Andhra Pradesh. In addition to the Oriya enumerated and localized above, the 300,000 Halba discussed below may be speaking a dialect of the same language. Oriya is also spoken as a second language by over 700,000 people.

Oriya is bounded by Bengali and Santali on the northeast, by Bihari on the north, by the Chhattisgarhi dialect of East Hindi on the northwest and west, and by Telegu on the south. The geographic boundary of Oriya is formed on the east by the Bay of Bengal, on the north roughly by the Haldi River, on the south by a line from Barwa to Tindiki, and on the west roughly by the Madhya Pradesh-Orissa border.

Though Oriya is centered in, and is the principal language of the state of Orissa, only 68% of Orissa's population speak Oriya as their native language. In the less densely populated areas of the state there are numerous tribes who speak Dravidian or Munda languages.

The most significant dialect variation is found in the fringe areas where Oriya dialects share features with the neighboring languages. The dialects of Oriya are:

Mughalbandi (Oriya Proper, Standard Oriya), spoken over all Orissa and even spilling over into Bihar, Madhya Pradesh and Andhra Pradesh, thereby overlapping the areas where the dialects listed below are spoken;

Southern Oriya (spoken by some of the lower castes in the district of Visakhapatam in Andhra Pradesh where Telegu is the dominant language) has incorporated features borrowed from Telegu;

Bhatri (spoken in the northeast of Bastar District of Madhya Pradesh) forms an intermediary speech form between the other Oriya dialects and Halbi, with which it shares many features;

Oriya (of the western and northwestern Orissa—Madhya Pradesh borderland, spoken in Sombalpur, Sundergarh, and Kalahandi Districts of Orissa and in Raipur and Raigarh Districts of Madhya Pradesh) shares many features with the Chattisgarh dialect of Eastern Hindi which is spoken in the same districts of Madhya Pradesh;

Northwest Oriya (spoken in the northernmost part of Raigarh District) shares features both with the Chhattisgarhi dialect of Eastern Hindi and with the Wagpuria dialect of Bihari;

Oriya of North Balasore (spoken in the northern Balasore District of Orissa) shares a great many features with Bengali; Grierson reports that it 'is non-intelligible to a speaker of Oriya from Puri and vice versa'; whether or not there is neighboring intelligibility between Oriya of North Balasore and the Oriya of the surrounding districts is not explicitly stated by Grierson, but the implication seems to be that there is.

Oriya of Midnapore (spoken in the Midnapore District of West Bengal) shares a great many features with Bengali and (in the northwest of Midnapore) has also borrowed from the Munda Santali;

Halbi (Halabi, Mahari, Mehari) is spoken by about 300,000 Halba, 87% of whom live in the Madhya Pradesh District of Bastar. The majority of the remaining 13% live in the Balaghat District of the same state, although a few are found distributed in Chanda, Bhandara, and other adjacent districts. Across the border in Orissa, Koraput District, as well as in other districts of Orissa adjacent to Madhya Pradesh State, small minorities of Halbi are found.

The status of Halbi as a dialect of Oriya is uncertain. Grierson treated Halbi as a dialect of Marathi 'for convenience', but noted that it shared a great many features with the Bhatri dialect of Oriya also. A mediary or transitional status for Halbi was claimed by Grierson when he stated that a person may travel nearly 1,500 miles from the easternmost Marathi region (the Arabian Sea Coast) westward through Halbi into Oriya, thence through Bengali to the westernmost boundary of Assamese without being able to establish sharp linguistic boundaries. The problem is complicated by the fact that where language barriers do exist, extensive language contact has caused diffusion of certain features across language boundaries. And Grierson does not specify whether the continuity he discusses represents merely a continuity in the occurrence of linguistic features (i.e. the lack of clusters of isoglosses) —without reference to mutual intelligibility between dialects—or whether the continuity does indeed represent mutual intelligibility between neighboring dialects. Some Indic linguists now follow Sten Konow in classing Halbi as a dialect of Oriya and more recent investigators have indicated that Marathi has definite linguistic boundaries which would fit with the assignment of Halbi to Oriya. However, Chatterji, among others, continues to list Halbi as a dialect of Marathi.

Halbi is by no means uniform; it is differentiated into a number of subdialects including:

Adkuri;
Bastari;
Chandari;
Gachikolo;
Mehari;
Muri;
Sundi.

The subdialects listed above form a homogeneous subgroup as opposed to the following more divergent subdialects:

Bunjia, spoken especially in Raipur, but also in Hoshangabad, Sambalpur, and Kalshandi Districts of Mathya Pradesh and Orissa;

Nahari, spoken in the districts of Raipur, Bilaspur, and Sambalpur; and in addition, in districts of the Chhattisgarh Division of Madhya Pradesh and Orissa;

Kamari (Kawari), spoken in Raipur and surrounding districts of Madhya Pradesh.

NORTHWEST ZONE

The two languages of the Northwest Zone are spoken by about 20 million people altogether, of whom three-fourths are Lahnda speakers. Lahnda and Sindhi are spoken chiefly in West Pakistan between the Indo-Pakistan border and the Indus River, and in the plains regions beyond the Indus to a point just east of Quetta. On a north-south axis they are spoken from Jammu and Kashmir in the north to the Kutch District of Gujarat in India.

(8) Lahnda (Western Panjabi, Hindki) is spoken by about 15,000,000 people in northeast Pakistan, west of the Indian States of Rajasthan and Punjab, southwest of Jammu and Kashmir, and in Rawalpind, Multan, and parts of the States of Bahawalpur, Dera Ismail Khan, and Peshawar in Pakistan. It is spoken southwest of the Dardic languages, east of the Iranian languages and the Dravidian Brahui, and west of Panjabi and Rajasthani.

The dialects of Lahnda according to Grierson are:

Standard Lahnda (Southern Lahnda), spoken throughout the Panjabi area of West Pakistan south of the Salt Range, has three subdialects:

Standard Lahnda, spoken in the districts of Shahpur, Jhang, Lyallpur, Montgomery, Gujranwala, and Gujarat;

Multani, spoken in the districts of Multan, Muzaffargarh, Dera Ghazi Khan, in the north of the State of Bahawalpur, and by numerous immigrants in Sind;

Thali, spoken in an area bordered by the Salt Range on the north, the Indus on the west, and the district of Muzaffargarh on the south, including parts of the Districts of Mianwali, Jhelum (Jehlam), Shahpur, Jhang, and the north of Muzaffargarh;

Northwestern Lahnda, spoken in an area beginning in the middle of the Salt Range, bordered by Thali on the south and by Northeastern Lahnda on the east and west, and extending northward through Western Jhelum into the eastern part of the Attack District, across North Attack and into the Hazara District, and westward across Peshawar, where it is spoken only by scattered Hindus;

Northeastern Lahnda is spoken in the eastern third of the Salt Range and the Pothwar Plain in the eastern part of the District of Jhelum and the plains portion of the District of Rawalpindi; and northward into the Murree Hills of Rawalpindi and a small tract in the east of the District of Hazara. Northeastern Lahnda is also spoken in the Chibhal country or outer hill region of Kashmir between the rivers Chenab and Jehlam, including the Jagir of Punch; and in the Jehlam Valley in Kashmir between Muzaffarabad and Uri, and the greater part of the valley of the Kishanganga; and in the western third of the Salt Range in the north of the District of Shahpur (separated from other Northeastern Lahnda by Northwestern Lahnda); and in the south and center of Attack, and across the Indus in Kohat.

(9) Sindhi is spoken by about 5 million people in the states of Gujarat and Rajasthan in India, and Hyderabad, Kharpur, Kalat and Quetta in Pakistan. It is bounded to the northeast by Lahnda, to the east by Rajasthani, to the southeast by Gujarati, to the west by Brahui (Dravidian) and Balochi (Iranian), and to the northwest by Pashto (Iranian). Large displacements of Lahnda and Sindhi people have taken place since the partition of India and

Pakistan. From the 1951 censuses of India and Pakistan it appears that less than a million Sindhi now reside in India; but the majority of Sindhi are still located in West Pakistan.

Grierson lists the following dialects of Sindhi:

Kachchi, spoken by hundreds of thousands of people previously located for the most part in the Kutch District of Gujarat State, but probably now largely displaced into Pakistan;

Thareli, spoken by several hundred thousand people along the western border of Rajasthan State in India, and in the western portions of Hyderabad and Khaipur in West Pakistan;

Vicholo (Central Sindhi), spoken by more than two million people in central and northern Hyderabad and southern Khaipur in West Pakistan;

Lari, spoken by tens of thousands of people in the northwestern portions of Kutch District in India, and the southern portion of Hyderabad in Pakistan;

Siraiki, perhaps merely a variety of Vicholo, spoken north of Vicholo in northern Khairpur to the junction of Khairpur, Quetta, and Kalat, and in a small section of northeastern Kalat;

Lasi, spoken by tens of thousands of people living west of the speakers of the Vicholo subdialect and north of, and west along the coast from, the city of Karachi in the district of Las Bela, Kalat;

Macharia, spoken in the Kapurthala District of the State of Panjab by a migratory tribe of fowlers (called 'gypsies' by Grierson), whose speech is Sindhi with a number of loans from Panjabi.

THE SOUTHERN ZONE OR MARATHI

(10) Marathi is the native language of about 28.5 million people (1951 census); it is the predominant language of Maharashtra State, where 85% of the total number of native Marathi speakers reside, while almost four million Marathi speakers live outside of Maharashtra. In addition to the speakers enumerated above, the 1951 census reports two and a half million persons who speak Marathi as a second language. Most native speakers of Marathi living outside of Maharashtra live in adjacent states. More than a million are reported to live in Madhya Pradesh, another million in Mysore, 100,000 in Andhra Pradesh. The remaining speakers of Marathi are found in almost every other state, but not more than 20,000 speakers live in any one non-adjacent state.

Between 1891 (when there were 18 million Marathi speakers) and 1951, there has been a 58% increase. Four districts of Maharashtra State, for which Grierson listed no speakers of Marathi in 1891, are reported to have altogether three million Marathi speakers in 1951—the districts are East Khandesh, West Khandesh, Surat, and Dangs. Otherwise the Marathi locales have changed little since 1891. In the listing of the main dialects which follows, the 1951 census figures are given for the districts in which Grierson located the dialects. The number of Marathi speakers in East and West Khandesh are included in the figure for the Deccan Marathi dialect, and the speakers in Dangs and Surat Districts are included in the figure for the Konkan Standard dialect.

Linguistically, Marathi is bounded by Gujarati on the northwest; by the Bhili dialects, Rajasthani, and Western Hindi on the north; by Eastern Hindi, the Halbi dialect of Oriya and the Dravidian Gondi on the east; and by the Dravidian languages Kanada and Telegu on the south. Ernest Bender reports that 'Marathi ... possesses a distinct linguistic frontier ... partly due to the barrier of the Vindhya Range'. However the Vindhya Range serves only to separate Marathi from Gujarati and does not eliminate the possibility of transitional dialects between Marathi and languages bordering it in other directions, as Halbi, which is sometimes listed as a dialect of Marathi and sometimes listed as a dialect of Oriya. Dialects which are transitional between Marathi and Eastern Hindi have also been reported.

Grierson notes that Marathi 'is a remarkably uniform language' with 'comparatively small ... dialect differences'. Nevertheless, Grierson identifies four main dialects of Marathi, discussed below in west to east order:

Konkan Standard (spoken in the coastal strip of North and Central Konkan);
Konkani (in the southern portion of the Konkan coastal strip);
Deccan Marathi (in central Maharashtra);
Varhadi-Nagpuri Marathi (in eastern Maharashtra).

Konkan Standard (Bankoti, Kunabi, North and Central Konkan) is spoken in the north and central sections of the Konkan coastal strip of Maharashtra bounded on the east by the western Ghats where the dialect merges with the Deccan dialect, on the north by Gujarati and on the south by the Konkani dialect. Approximately three million people (a little under 10% of Marathi speakers) speak the Konkan Standard dialect. Grierson lists the following subdialects of Konkan Standard:

Parabhi (Kayasthi, Damani), spoken by the Marathi speaking population of Bombay and Thana Districts as far north as Daman;

Koli, spoken by small tribes called the Hill Koli, the Son Koli of the coast, and 'certain low-castes ... not recognized by the rest' in the districts of Bombay, Thana, Kolaba, and Janjira;

Kiristãv, spoken by the native Christians of Thana District;

Agari of Kolaba, spoken in Kolaba District by 'a class of husbandmen who inhabit villages on both sides of the Amba River in the Pen and Alibak Talukas, in villages situated on the creeks of the Panwel Taluka, and in all villages of the Uran Peta';

Dhanagari, spoken by the shepherds of the districts of Thana, Jawhar, Janjira, Belgaum;

Bhandari, spoken by the palm-juice drawers (the Bhandaris);

Thakari, spoken by the Thakars who are distributed extensively among the Gujarati speakers in Gujarat and among the speakers of Deccan and other Konkan Marathi;

Karhadi, spoken by the Karhada Brahmans in Sawantwadi, Ratnagiri, and Bombay Districts;

Sangamesvari (Bakoti, Bankoti), spoken by Marathi speakers in the region from Bombay to Rajapur and also to the south of Kolaba by the Mohammedans whose speech is called Bakoti;

Ghati, spoken in the western Ghats between Kolaba and Bhor Districts, is 'probably identical with Maoli', the language of Maval, the country above Sahyadris, between Thana and Poona Districts;

Mahari (Dhed, Holia, Parvari), spoken by the Mahars distributed over the area of Bombay Districts.

The subdialects listed above are closely interrelated; the following subdialects are linguistically more differentiated from each other but share many features with languages of the Central Zone:

Katkari (Kathodi, Katvadi); spoken by a forest tribe in Konkan and Sahyadri Hills, Katkari shares a great many features with Khandesi;

Varli (spoken in the northwest in Dahanu, Mokhada, Murbad, Kalyan, and Karjat of the Thana District, in the districts of Jawhar, and in the Satputra Range in Khandesh) is sometimes classified as a subdialect of Gujarati or Bhili because it shares a great many features with them;

Vadval (Phudagi); spoken in the coastal subdistricts of Bassein, Mahim, and Dahanu of the Thana District, Vadval is more like the less divergent group of subdialects listed above than are the other subdialects of this list; nevertheless, it shares 'some phonetical changes' with languages of the Central Zone;

Phudagi, very closely related to Vadval above, is spoken by a wandering tribe in Thana District;

Samvedi (spoken by a group of husbandmen and gardeners in Bassein and

Mahim) shares many features with Gujarati;

Mangelas, a minor caste subdialect in the north of Thana District, also shares a great many features with Gujarati.

Konkani (Gomataki, Goanese) is the dialect of Marathi spoken on the southern Konkan coastal strip of Maharashtra, primarily in the district of Ratnagiri and the area of Goa (recently annexed from Portugal). Konkani is also spoken in the Mysore District of Ramra Shimgoa, and the south Kamara District of Kerala State—altogether by two million people (7% of the total number of Marathi speakers). Konkani stands apart from the Konkan Standard, Deccan and Varhadi-Nagpuri dialects, which together form a closely interrelated group in the subgrouping of Marathi dialects. Between the southernmost subdialects of Konkan Standard (Sangamesvari and Bankoti) and the northernmost subdialect of Konkani (Kudali), there are two transitional subdialects, Daldi and Chitpavani, which form 'connecting links' between the two dialects according to Grierson. The subdialects of Konkani, which vary 'according to locality and to the caste of speakers', listed by Grierson are:

Standard Konkani, spoken in north Kanara, Goa, Sawantwadi, and Bombay Districts and also in Vengurla and Mahwan of the southern corner of Ratnagiri District;

Bardeskari (Gomantaki), spoken in the southwest of Belgaum District of Mysore;

Sarasvat Brahman subdialect, spoken by large numbers of Sarasvat Brahmans

in the towns and villages of Karwar and Ankola on the coast; and inland in Haliyal, Supa and Sirsi;
Kudali (Malvani), spoken in Sawantwadi and the southern part of Ratnagiri District from the Santarda River on the south to Deogad, Kankoli, and Phanda Ghat in the north.

The two subdialects which follow have more features in common with the Konkan Standard dialect than do the other subdialects of Konkani:
Daldi (Nawaits), spoken by Mohammedan dishermen located in districts adjacent to Mysore, and the districts of Ratnagiri, Janjira, and Bombay;
Chitpavani (Konkanasths), spoken by the chief Konkan Brahmans in Ratnagiri and Bombay Districts, and in the Savantvade towns of Vadi, Kudal, Banda, and in the villages near Sahyadris, and in the Ajgaons subdivision.

The Deccan (Desi, Dakini) dialect is spoken by an estimated 14,500,000 people (51% of the total number of Marathi speakers), in the central section of Maharashtra extending east from the Konkan coastal strip (where Konkan Standard and Konkani are spoken) to the region where the Varhadi-Nagpuri dialect begins, in an area roughly corresponding to the districts of Buldana, Akola, and Yeotmal. South of Maharashtra, the Deccan dialect extends into the districts of Belgaum and Bijapur in Mysore; northwest of Maharashtra, the Deccan dialect is spoken in the Madhya Pradesh Districts of Raisen and Sehore, and in Gujarat State. It is spoken by the Brahmans and members of other high castes in the Thana Districts from Daman to Rajapur where the Konkan Standard dialect predominates. In addition to the Deccan proper subdialect, there are two other subdialects of the Deccan dialect. The Deccan Marathi spoken in the Western Ghats shares many features with Konkan Standard as opposed to Deccan proper, but is not listed as a separate subdialect by Grierson, as are:

Kalvadi (Dharwar), spoken by the Kunabi in Dharwar and Kanara Districts of Mysore;

Bijapuri, spoken by rural people of the Bijapur District of Mysore.

The Varhadi-Nagpuri dialect of Marathi (Madhya Pradesh Marathi, Berari, Marathi of Central Provinces and Berar, Dhanagari, Kumbhari) is spoken in the state of Maharashtra by over five million people (18% of the total number of Marathi speakers). The Varhadi-Nagpuri dialect is also spoken by 236,000 people in the Chhindwara and Balaghat Districts of Madhya Pradesh, and by well over 200,000 in Adilabad and Nizamabad Districts of Andhra Pradesh. Subdialects of the Varhadi-Nagpuri dialect are:

Brahmani, a Varhadi subdialect spoken by 18,000 'educated people' in Akola District and the eastern area of Buldana District in Berar;

Kunbi, a Varhadi subdialect spoken by 443,600 uneducated people of the Akola District of Berar;

Raipur, a Nagpuri subdialect spoken in villages to the south of Nandgoon;

Dzharpi (Jhadpi), spoken by 5,000 people in 1891 in the Ellichpur District of Berar;

Govari of Bhandara, spoken by 150 cowherds of Bhandara;

Kosti (Rangari), spoken by approximately 3,000 weavers and dyers of

Berar in Akola, Ellichpur, and Buldana Districts;

Kunban (Kohli), spoken in Chanda District by 110,150 husbandmen and rice-growing, tank-making cultivators in 1891;

Mahari (Dhedi), spoken by 19,000 village-watchmen, gate-keepers, messengers, guides, porters, and the like in the Chhindwara District.

The above dialects form a coordinate group, Varhadi-Nagpuri proper; the following subdialects are more divergent:

Marheti, spoken in the southern part of the Balaghat District of Madhya Pradesh by the 'lower class' of that district;

Natakani, spoken in 1891 by 180 people in the Sironcha of the Chanda District of Maharashtra;

Katia (Katiyai); spoken by weavers and village watchmen in Maharashtra and in the Chhindwara and Hoshangabad Districts of Madhya Pradesh, Katia shares a great many features with East Hindi.

CENTRAL ZONE

The Central Zone Indic languages are spoken by a total of 125 million people in an area of western India south of the Himalayas between the Pakistan border and the center of the Indian subcontinent. Almost half of the 120 million people in the area speak Western Hindi. Half of the remaining half dozen languages in the Central Zone are spoken by almost 20 million people each—Rajasthani, Panjabi, Gujarati.

(11) Gujarati (Gujerati) is the native language of inhabitants of the State of Gujarat except in the district of Kutch, where Gujarati is known not as a native language but as a lingua franca. It is also spoken by about a tenth of the people of the adjacent district of Thana of the State of Maharashtra on the south; and on the north there are a considerable number of Gujarati speakers in Hyderabad State in Pakistan. There are also large numbers of Gujarati speakers in other states, particularly in urban centers.

In 1951 more than sixteen million spoke Gujarati as a native language and another half million spoke it as a second language. 15,150,000 of the native speakers of Gujarati lived in the State of Gujarat in 1951. The remaining 1,150,000 were living in the following states: Bombay (840,000), Madhya Pradesh (160,000), Rajasthan (53,000), Mysore (26,000), Andhra Pradesh (16,000), West Bengal (15,000), Madras (15,000), Punjab and Delhi (14,000), Uttar Pradesh (14,000); less than 10,000 each in Bihar, Orissa, and Kerala; less than 500 each in the Andaman and Nicobar Islands and Assam.

The vast majority of Gujarati speakers are monolingual; only a little more than three percent (475,000) are bilingual. Of these, 445,000 are bilingual in other Indic languages: 280,000 bilingual in Hindi-Urdu; 130,000 bilingual in neighboring Marathi; 16,000 in neighboring Ranjasthani. Another 27,000 speak Dravidian languages; most are bilingual in Kannada (18,000).

Grierson subdivides Gujarati into 'educated' and 'uneducated' dialects. It appears, however, from his description that there are several varieties of both 'educated'and 'uneducated'. The following classification and discussion is based on Grierson's dialect names and his comments on them. Grierson gives numbers of speakers for some localities—but none which distinguish the number of speakers of Standard Gujarati from the number of speakers of another dialect in a particular area.

Standard Gujarati is the term used to cover the speech of the educated people in every region where Gujarati is spoken; hence there is in fact more than one Standard. The following four subdialects are distinguished:

Saurashtra Standard, the speech of educated speakers of the State of Gujarati;

Nagari, the speech of Nagar Brahmans, includes more Sanskrit words than other dialects;

Bombay Gujarati (spoken by educated speakers in and around Bombay) includes Marathi loan words;

Patnuli (Saurashtri), spoken by the silk weavers of Madras and the Deccan.

Gamadia (Gramya) is a general term for the dialect of uneducated village people. Gamadia is distinguished from Standard Gujarati by tendencies to drop /h/, replace /s/ with /h/, replace /čh/ with /s/, metathesize /y/ and a following

consonant, and to lose the contrast between dental and retroflexed consonants. Eight subdialects appear to be classifiable as belonging to Gamadia:
Surati, spoken in Surat and Broach;
Anawla (Bhathela), spoken by Bhathela or Anawla Brahmans of Surat, Baroda, and other areas bordering on the Bhili dialects;
Eastern Broach Gujarati (spoken in the eastern part of Broach) shows considerable affiliation with the Bhili dialects;
Charotari, spoken in part of the district of Kaira and parts of Baroda;
Patidari, spoken in part of the district of Kaira;
Vadodari, spoken in the district of Baroda and part of Kaira;
Ahemedabad Gamadia, spoken in the villages of central and northwest Ahemedabad District;
Patani, spoken in parts of eastern Mehsana District, in Banaskantha District, and in Sabarkantha District. It is very likely that the Gujarati spoken in Pakistan is also of the Patani subdialect.

Parsi (the dialect spoken by the Parsis) is characterized by more Arabic and Persian borrowings than the other dialects.

Kathiyawadi is the dialect spoken on the Peninsula of Kathiawar. There were an estimated 2,600,000 speakers in 1901. Kathiyawadi is divided into four subdialects:
Jhalawadi, spoken in the northeastern portion of the peninsula by about 440,000 people in 1901;
Sorathi, spoken in the southwestern portion by 730,000 people in 1901;

Holadi, spoken in the central and northwestern portions by 770,000 people in 1901;

Gohilwadi (Bhawnagari), spoken in the southeast by 630,000 people in 1901.

Kharwa is the dialect of Gujarati spoken by Moslem seamen on the island of Piram and in the part of Ahmedabad District on the east coast of Kathiawar Peninsula. The grammar of Kharwa is essentially the same as that of Standard Gujarati; the phonology is strikingly different.

Kakari is listed by Grierson as a 'mixed'dialect, based mainly on Gujarati. It was spoken by 122 Kakars in the Deccan of Peninsular India in 1891. Apparently the other 34,000 Kakars did not speak Kakari but the languages of the areas in which they settled.

Tarimuki (Ghisadi) is the dialect of wandering blacksmiths in Poona, Satara, Belgaum, Amrasti, Akola, Buldana, and perhaps other cities in southern India. There were 1,700 known speakers in 1901. Their dialect is fairly similar to the Gamadia subdialects.

(12) Rajasthani is widely spoken by twenty-two million people in most of the districts of Rajasthan (excluding, however, the Bharatpur District, all but the western fourth of Sawai Madhopur District, the eastern tip of Jaipur, the northern third of Ganganagar, the western fourth of Jaisalmer, and a strip inside the western and southwestern borders of Udaipur, and all of Bungarpur and Banswara). Rajasthani is also spoken along the eastern border of Bahawalpur in Pakistan; north into the Panjab: in the Gurgaon District; in all but the northern tip of Mohindergarh, and along the southwestern border and through the central part

of Hissar; in Madhya Pradesh State in the districts of Mansaur, Ratlam, Ujjain, Shajapur, Rajgarh, the northern halves of Dhar, Indove, Dewas, Raisen; in all but the eastern tip of Sehore, Betul, eastern Nimar, Goona Bhilsa, the western edge of Shivpuri, and the western third of Morena; and there is an isolated Rajasthani enclave in western Nimar. Rajasthani is bordered on the east by Western Hindi, on the north by Panjabi, on the northwest by Lahnda, on the west by Sindhi, on the southeast by Bhili, and on the south by Gujarati and Marathi.

In the area in Madhya Pradesh that was inhabited by Rajasthani speakers at the time of Grierson's survey, the 1951 census reports less than a million speakers for Rajasthani but over four million 'Hindi' speakers. This could mean either that Western Hindi has become the major language of Madhya Pradesh or that the speakers of the Malvi dialect of Rajasthani—intermediate between Western Hindi and Rajasthani and Gujarati dialects—reported their language to be Hindi in the language census. Or both factors may be involved. Similarly, in the Gurgaon and Mohindergarh Districts of the Panjab and the Alwar District of Rajasthan, Grierson describes the dialect of Mewati as transitional between Rajasthani and Western Hindi, but the 1951 census reports over two million Hindi and less than 200,000 speakers of Rajasthani dialects (all of them in Alwar).

Rajasthani is differentiated into five dialects:

Marwari (Merwari, Mewari) is the dialect of the Rajasthani area, spoken by nine million people in all parts of Rajasthan except the districts of Jaipur, Kotah, Tonk, Alwar, Bundi Jhalawar and eastern Ajmer. Marwari is also spoken in Pakistan and north into the Hissar District of the Panjab. The Marwari dialect has five subdialects:

Standard Marwari, spoken in north, south, and east Nagore, northwest Ajmer, the northern half of Pali, southeast Jodhpur, southeast Barmer, northeastern Jalore, northwestern Sirohi;

Eastern Marwari, spoken in eastern Ajmer, the southern extension of Jaipur, the southern half of Tonk, all of Bundi, Bhilwara and Chittorgarh, and in eastern Udaipur;

Southern Marwari (Marwari-Gujarati), spoken in western Udaipur, southern Pali, southern and eastern Sirohi, and southern Jalore;

Western Marwari, spoken in north, east, central, and south Jaisalmer, western and north Barmer, extreme western Jalore, north and west Jodhpur, and northeast Nagore, and also in the adjacent areas of Pakistan;

Northern Marwari, spoken in Ganganagar, Bikaner, Churu, Jhunjhunu, and Sikar Districts of Rajasthan, and also in the Hissar District of Punjab State.

The Central-eastern dialect of Rajasthani is spoken by some four million people in the following districts: Jaipur, eastern Sikar, northern Tonk, western Sawai Madhopur, Kotah, and northern Jhalawar, and eastern Bundi. It is also spoken in Morena and Shivpuri Districts of Madhya Bharat Province. It has four subdialects:

Ajmeri, spoken in western Jaipur and northwestern Tonk;

Kishangari, spoken in a strip just to the east of Ajmeri in the same districts as Ajmeri;

Havauti, spoken in Kotah, northern Jhalawar, and eastern Bundi;

Jaipuri, spoken in eastern Sikar, Jaipur, northeastern Tonk, and western Sawai Madhopur (with a divergent form, Ladi, spoken by a wandering tribe of vendors primarily located in the state of Maharashtra).

The Northeastern dialect of Rajasthani is spoken in the district of Alwar, and in the northern parts of Jaipur, Bharatpur and Sawai Madhopur. It is also spoken in Gurgaon and Mohindergarh Districts of the State of Punjab. According to Grierson, the northeastern dialect has two subdialects:

Mewati, spoken in the districts of Alwar, and northern Jaipur, Bharatpur and Sawai Madhopur;

Ahirwati, spoken in Gurgaon and Mohindergarh Districts of Punjab Province.

Malvi (Malavi) is the dialect of Rajasthani spoken in Chittorgarh and southwestern Jhalawar Districts. It is also spoken in Rajgarh, Shajapur, Mandsaur, Ratlam, Ujjain, western Bhilsa, and western Goona Districts of Madhya Bharat Province, in Raisen and northwest Sehore Districts of Bhopal Province, and in Hoshangabad, Betul, and Nimar Districts of Madhya Pradesh Province.

There are two subdialects:

Sondwari, spoken in southwestern Jhalawar;

Malvi, spoken in the rest of the Malvi dialect area.

Nimadi is the dialect of Rajasthani spoken in small enclaves in extreme western Hoshangabad, northeast Nimar, and southern Dewas Districts in Madhya Pradesh.

Gujuri is the Rajasthani dialect spoken by the Gujurs of the hills north of the Punjab in the Pahari area. According to Grierson, Gujuri is especially closely related to the Mewati and Marwari dialects as against the intervening Jaipuri dialect. Gujuri is spoken in three areas: Hazara, Swat, and Kashmir.

Among the Central Zone Indic languages, Gujarati and Rajasthani, (11) and (12) above, bear an especially close relationship to each other and to the three languages listed below—(13) Bhili, (14) Khandesi, and (15) Banjuri. The exact nature and the degree of closeness of the relationship is not known.

Bhili, Khandesi and Banjuri—listed immediately below—are particularly conservative in respect to the retention of earlier Indic features, but all have many Western Hindi, Marathi, Gujarathi, and Rajasthani loanwords. Bhili and Khandesi are described by Grierson as standing linguistically between Gujarati and Rajasthani. Banjuri, however, seems somewhat closer to Rajasthani.

(13) Bhili is spoken by less than two million people in the mountainous area where the States of Rajasthan, Madhya Pradesh, Gujarat, and Maharashtra converge. It is spoken in the southern part of Rajasthan in the districts of Udaipur, Dungarpur, Banswara and Chittorgarh; in the western part of Madhya Pradesh in the districts of Ratlam, Thabua, Dhar and Nimar; in the eastern and south-eastern part of Gujarat in the districts of Baroda, Amreli and Surat. It is spoken primarily in the northern districts of West and East Kandesh and south into Aurangabad in Maharashtra. Peripheral to this area smaller numbers of Bhili speakers are scattered, but in diminishing numbers as Bhili is superseded as the language of many by the larger neighboring languages Rajasthani, Gujarati,

Marathi, and Western Hindi. Though only about a million Bhili speakers were enumerated in the 1951 census, many other bilingual Bhili speakers probably reported themselves as speaking Hindi, Gujarati, Marathi, or Rajasthani. Areas which formerly contained large numbers of Bhili speakers, as the Kutch District of Gujarat, now claim few, if any, Bhili speakers.

Grierson treats Bhili as forming a southern link between Gujarati and Rajasthani, but says that Bhili could be considered an eastern dialect of Gujarati, though it is'almost identical with' Khandesi, which he regards as a possible dialect of Rajasthani.

Though Grierson indicated that dialectal diversity in Bhili seems slight, he enumerated twenty-eight distinct variants of Bhili:

Siyalgiri, spoken by a 'wandering tribe' in Orissa numbering in the hundreds;
Baori, spoken by a group in the Punjab numbering nearly 50,000;
Ahiri
Anarya (Pahadi)
Barei
Charani
Chodhari
Dehawali
Dhodia
Dubli
Gamati
Girasia
Habura

Konkani

Kotali

Magara ki Boli

Mawchi

Nahari (Baglani)

Naikadi

Panchali

Pardhi (Takankari)

Pawari

Ranawat

Rani Bhil

Rathavi

Wagadi.

(14) Khandesi (Ahirani, Dhed Gujari) is spoken by about a million people in East and West Khandesi, and in the neighboring districts of Nisik in Maharashtra State, and Akola, Nimar and Buldana in Madhya Pradesh. There are four subdialects:

Khandesi proper

Dangri

Rangari

Kunbi (Kunbau).

(15) The Banjuri (Labhani, Bahrupia) are a tribe of 'carriers' found in western and Southern India. Banjuri resembles northern Gujarati, but according to Grierson, must ultimately be referred to as Rajasthani.

There are two principal dialects of Banjuri: that of the Panjab and Gujarat, and that spoken in other areas of which the dialect spoken by the Labhana of Berar is the Standard. To these are added:

Labanki, in Muzaffargarh District of the Punjab;
Kakeri, in the Jhansi District of Uttar Pradesh;
Bahrupia, in the Punjab.

(16) Panjabi (Eastern Panjabi, Punjabi) is spoken by some twenty million people in northwestern India (in the State of Punjab, the Ganganaga District of Rajasthan) and across the border in Pakistan to just west of the city of Lahore. Panjabi is bounded to the northeast by Pahari in the lower ranges of the Himalayas, to the north by Dardic languages, to the south by Rajasthani, to the west by Lahnda. But in the west, beyond the city of Lahore in Pakistan, there appears to be partial intelligibility with dialects of the Lahnda language. The southeastern and eastern boundaries are also difficult to establish. Panjabi is bordered to the east by Western Hindi; a convenient dividing line between them may be drawn north and south along the Ghaggar River on about the northern boundary of the Karnal District and bisecting the Sangur and Ambola Districts in a northeast to southwest direction. Concerning the boundary between Panjabi and Urdu (one particular form of the Hindustani dialect of Western Hindi), Hamid Ahmad Khan (The Common Structural Basis of Urdu and Panjabi, Pakistani Linguistics, 1963) says: "... it may be borne in mind that these two are not distinct languages in the sense that German and French are. In the case of French and German there is a certain frontier

line at which the one suddenly ceases and the other suddenly begins. Panjabi and Urdu have never been thus separated in terms of a geographical line. Right from Rawalpindi in Pakistan to Ambala in India, we have district after district speaking one dialect of Panjabi after another. As we go southeast [from Ambala] a continuous process of dialect change, not a sudden or violent break ... converts the Panjabi of Ambola to the Urdü of Saharanpur and Delhi." A similar situation exists in respect to the boundary between the Panjabi dialects of India and Rajasthani. Grierson indicates that there is a gradual merging into Rajasthani, through the intermediate Battiani subdialect of Panjabi proper.

There are two main dialect divisions of Panjabi, Panjabi proper and Dogri-Kongri.

Panjabi proper is spoken in India in the districts of Gurdaspur, Amritsar, Hoshiarpur, Jullundur, Kapurthala, the northern portion of Ferozpur, Ganganaga Ludhiana, Fatehgarh Sahib, Ambala, the northern portion of Roktak, Barnala, Bhatinda, and Sangur (all of these are in Panjab State except the Ganganaga District of Rajasthan), and in West Pakistan in the Lahore division. There are six main subdialects:

Majhi, spoken in Laphore (Pakistan), Gurdaspur and Amritsan (India);
Doab, spoken in Hosiarpur, Jallundar and Kapurthala;
Bhattiana, spoken in Ferozpur (southern portion), and Ganganaga (Rajasthan State) Districts;
Powadhi, spoken in Ludhiana, Fategarh Sahib, the northern portion of Ambala and Roktak Districts;

Malwa; spoken in Barnala, Bhatinda, the northern half of Sangur, and the northern portion of Ferozpur;
Bathi, spoken in the southern parts of the Bhatinda, Sangur, and Ruktak Districts.

The Dogri-Kangri dialect is spoken in the Kangra District of Panjab State, and the southern part of Jammu and Kuthua Districts of Jammu and Kashmir.

(17) Western Hindi (Hindi, Urdu, Hindustani, Hindusthani, Khari-boli) is the native language of some sixty million people in north central India, in an area which encompasses the central and western portions of the state of Uttar Pradesh, the northern part of the state of Madhya Pradesh, and the eastern part of the state of Punjab. In India Western Hindi is bounded on the northwest by Punjabi, on the west by Rajasthani, on the south by Marathi, on the east by Eastern Hindi, and on the northeast by the Pahari languages. There are also large colonies of Western Hindi native speakers in other parts of the world, particularly in the southern part of Africa where it is spoken by around two million people, and in the Guianas where 32 percent (100, 000) of the population of British Guiana and 49 percent (285, 000) of the population of Surinam (Dutch Guiana) speak Western Hindi. In addition, the literary forms of High Hindi and Urdu serve as literary languages for another 30 million people; and Bazaar Hindi (Chaltu, Chalu or Laghu Hindi) serves as a lingua franca for millions more.

Five dialects of Western Hindi are recognized:

Hindustani, spoken by nearly thirty million people;

Bangaru, spoken by nearly four million;

Braj Bhakha spoken by more than eleven and one half million;

Kanauji, spoken by approximately six million;

Bundeli, spoken by nearly eight million.

Hindustani is the native dialect of some thirty million speakers scattered throughout India. The area of Hindustani predominance includes the districts of Rampur, Moradabad, Bijnor, Meerut, Muzaffarnagar, Saharanpur, and the plains country in the southern part of Dehra Dun; all districts in Uttar Pradesh, and the eastern part of the Ambala District in the Punjab. According to the 1951 census at least half the population of Bareilly in Uttar Pradesh claim Hindustani as their native language, and the Hindustani dialect may have completely replaced the Braj Bhakha dialect in this district. Hindustani is also spoken by over ten percent of the population of the following districts outside of the area of Western Hindi predominance: Lucknow, Unnao, Kheri, Bahraich, Bara Banka, Mirzapur, and Allahbad in Uttar Pradesh; Cuddapah and Kurnool in Madras; Bangalore in Mysore; Greater Bombay and Dharwar in Maharashtra; Akola in Madhya Pradesh; Gulberga, Nizamabad, and Raichur in Andhra Pradesh; and on the Andaman and Nicobar Islands.

Although Hindustani is the native dialect of a relatively small fraction of India's total population, it has developed two literary forms, High Hindi and Urdu (together, Hindi-Urdu), which function as the official 'language' of India. Hindi-Urdu is also the literary vehicle for much of northern India;

but it is not the literary language in the states of Bengal, Assam, Orissa, Nepal, Gujarat, and Maharashtra. There are four varieties of this literary language; the first three listed below also represent spoken subdialects:

Hindi (High Hindi, Nagari Hindi, Literary Hindi, Standard Hindi), written in the Nagari script, which has been de-Persianized and de-Arabicized; many Sanskrit words have been introduced;

Urdu, written in Arabic script and containing many Persian and Arabic loans (11.5 million speakers);

Dakhini, also written in Arabic script, but freer of Persianization than Urdu (five million speakers);

Rekhta (Rekhti) a form of Urdu used in poetry.

Of the relationship of Hindi and Urdu, Chatterji (1945) says, "out of the same language grew two literary speeches, alien to each other in script and in higher vocabulary; and they started their rival careers as soon as they developed prose literatures, under English auspices in Calcutta from the very first decade of the 19th century, and began to be employed in schools and in public life. With the entry of those who spoke or used them into the field of politics, and with the ugly development of Hindu-Muslim communalism, High-Hindi and Urdu became symbols of this conflict. Each is going its own way; intense Persianization on the one hand and almost equally intense Sanskritization on the other. In their more 'elegant' forms one would be unintelligible to those who use the other."

The Bangaru (Hariani, Deswali, Desari, Jatu, Jati, Chamarwa) dialect of Western Hindi has close to four million speakers in the Panjab, the Rohtak, Karnal, and eastern part of Hissar, the southeast corner of Patiala, and in Delhi.

The Braj Bhakha (Brij Bhasha, Antarbedi, Antarvedi) dialect has over eleven and a half million speakers in the area including the districts of Bharatpur and Sawai Modhopur in Rajasthan and north through the districts of Manipuri, Agra, Etah, Mathura, Aligarh, Bulandshahr, Budaun, and possibly Bareilly; and along the southern edge of Nainital in Uttar Pradesh, as well as in a crescent along the upper part of the northwestern border in Morena in Madhya Pradesh, Grierson reports that the native speakers recognize seven subdialects:

Braj Bhakha proper, in nearly all of the area except the southern part;
Antarbedi, along the eastern edge of the northern part of the area;
Bhuksa, in southern Nainatal (sometimes mentioned as a subdialect of the Hindustani or Kanuji dialect);
Sikarwari, in the northern half of that part of the dialect area in Morena;
Jadobafi, in the southern half of that part of the dialect area in Morena, extending into Dangi country north in Sawai Madhopur;
Dangi, including the varieties Dugarwara, Kalimal, and Dangbhang, in Sawai Madhopur, spoken by 16,491 people in 1951; and there are East Agra subdialects spoken in eastern Agra District.

The Kanauji dialect is spoken by about six million people in the Pilibhit, Ferrukhabad, Shajahanpur, Hardoi, Etawah, and Kanpur Districts

6

of Uttar Pradesh. Kanauji has three subdialects:

Kanauji proper, spoken throughout most of the area;

Tirhari, spoken in the very southern part of Kanpur along the Jamna River (there is also a Tirhari subdialect of Eastern Hindi on the southern side of the river);

the mixed subdialect of the eastern tip of Hardoi, which is transitional between Kanauji and the Awadhi dialect of Eastern Hindi.

The Bundeli (Bundel Khandi) dialect has close to eight million speakers in Uttar Pradesh in the districts of Jalaun and Jhansi, and in the western three quarters of Hamirpur, and in Madhya Pradesh in the districts of Balaghat, Chhindwara, Hoshangabad, Sagar, the eastern edge of Sehore, the western edge of Satna, Panna, Chhatarpur, Tikangarh, Shirpuri, Gird, Bhind, and the eastern tip of Morena, plus scattered speakers to the south in the Bhandara and Nagpur Districts of Maharashtra.

Grierson suggests classifying Bundeli into the following seven subdialects:

Standard Braj of Mathura, Aligarh, and Western Agra;

Standard Braj of Bulandshahr;

Standard Braj of eastern Agra, the southern part of the area in Morena, and in southern Bharatpur;

Braj merging into Kanauji in Etah, Mainpuri, Budaun, and Bareilly;

Braj merging into the Bhadauri subdialect of Bundeli in the northern part of Morena;

Braj merging into Jaipuri (Rajasthani in Northern Bharatpur and in Sawai Uradhopur;

Rhuksa in southern Nainital,

The following thirteen subdialects, however, are recognized by the speakers of Bundeli;

Bundeli proper (Bundel Khandi), spoken throughout the greater portion of the area, including all of the south (except the southwestern part of Chhindwara), the western half of the area, and in most of the north central part of the area;

Pawari (Powari), spoken in Datia and the surrounding area; in the 1951 census 35,979 people, mostly in Balaghat, reported their language as 'Powari';

Lodhanti (Rathora), spoken around the northwest corner of Hamirpur;

Khatola, spoken throughout most of Panna and the western edge of Satna;

Banaphari, a transitional dialect between Bundeli and the Bagheli dialect of Eastern Hindi, spoken in the northern parts of Chhatarpur and Panna;

Kundri, a transitional dialect between Bundeli and the Bagheli dialect of Eastern Hindi;

Nibhatta, a transitional dialect between Bundeli and the Bagheli dialect of Eastern Hindi, spoken in a small area in eastern Jalaun;

Bhadauri (Towargarhi) is a transitional subdialect between Bundeli and the Braj Bhakha dialect of Western Hindi, spoken throughout Bhind, Gird, and Shirpuri;

Lodhi, listed as their native language by over 12,000 people in 1951, mostly in Balaghat;

Koshti, listed as their native language by over 10,500 people in 1951, mostly in Bhandara and Nagpur, as well as in Chhindwara;

Kumbhari, spoken in Chhindwara;

Nagpuri Hindi, spoken in Nagpur, south of the Bundeli area;

Chhindwara Bundeli, spoken in the southwestern part of Chhindwara.

EAST-CENTRAL ZONE OR EASTERN ZONE

(18) Eastern Hindi (Kosali) is spoken by some thirty million people in the western portions of the states of Uttar Pradesh and Madhya Pradesh. In Uttar Pradesh, Eastern Hindi is spoken in the entire Faizabad Division; in all of the Lucknow Division except the district of Hardai; in Fatehpur and Allahabad Districts in the Allahabad Division; and in the Banda District in Jansi Division. In Madhya Pradesh, Eastern Hindi is spoken in the former Baghel Khand Division in the now defunct Vindhya Pradesh (which has become the Northwest Madhya Pradesh Division); and in the Chhattisgarh plain region of the East Madhya Pradesh Division. The entire area of Eastern Hindi forms a narrow corridor, 250 miles wide at its broadest point, but 750 miles long, from the Indian-Nepalese border in the north into the district of Bastar in Madhya Pradesh in the south. The speakers are concentrated in the Uttar Pradesh area (14,500,000), with fewer speakers scattered through Madhya Pradesh (9,500,000). In addition to the speakers in this area, another million (speaking the Awadhi dialect of Eastern Hindi) were reported by Grierson to be scattered through the Bihari-speaking area in the westernmost districts of Uttar Pradesh, and in the easternmost districts of Bihar. Still another million speakers of Eastern Hindi are found in the

adjacent area of Nepal. Eastern Hindi is bounded to the north by Eastern Pahari (Nepali), to the east by Bihari, to the west by Western Hindi, to the southwest by Marathi, and to the southeast by Oriya.

The dialects listed below are those listed by Grierson.

The Awadhi-Bagheli dialect is spoken by about 20,750,000 people located in the Eastern Hindi area of Uttar Pradesh, a portion of the North-west Madhya Pradesh Division, and the districts of Jabalpur and Mandla in the East Madhya Pradesh Division. Awadhi speakers are also located in the easternmost portions of Uttar Pradesh, the adjacent areas of Bihar, and in the contiguous areas of Nepal.

Awadhi and Bagheli constitute two subdialects of the same dialect. Awadhi (Kosali, Baiswari) is spoken by about 16 million people in the districts of Kheri, Sitapur, Lucknow, Unao, Rae-Bareli, Bahraich, Bara Banki, Pratapagarh, Sultanpur, Gonda, Faizbad, and Allahabad in Uttar Pradesh. Awadhi is the standard literary variety of Eastern Hindi, with a literature dating from the 16th century.

Bagheli (Baghelkhandi, Riwai) is spoken by about 4,750,000 people in the Banda District of Uttar Pradesh (and extending to the north bank of the Jamni River in the Fatehpur District), and in the Satna, Rewa, Shadal, Sidhi, Jabalpur, and Mandla Districts of Madhya Pradesh. In the Banda District and adjacent areas a great deal of dialect differentiation has taken place, complicated in some cases by the borrowing of West Hindi features resulting in a number of varieties (e.g. Tirhari, Gahore, Banaphari). Similarly in the Mandla District and environs there are a number of divergent varieties, as Marari and Powari, and Ojhi (in Chindwara District of Madhya Pradesh). There are also a couple of other regional variations mentioned—Godwani (Mandlaha) of Mandla District; and Sonpari of the Shadhol District. Bagheli has had a literature since the 16th century.

The Chhattisgarhi (Laria, Khatahi) dialect of Eastern Hindi is spoken by about six million people on the Chhattisgarh plain of Madhya Pradesh in the districts of Bilaspur, Surguja, Raigarh, Raipur, Durg; and, to a lesser extent, in Bastar, Balaghat, Sambalalpur (in Orissa State). A few are also reported in the Patna District of Bihar State. Grierson lists the following subdialects:

Surgujia, spoken in the districts of Surguja and Raigarh;
Sadri Korwa, spoken by the Korwa tribe in Jashpur District;
Baigani, in Balaghat, Raipur, Bilaspur, and Sambalalpur;
Binjhwari, in Raipur, Raigarh, and Patna, in Bihar;
Kalanga, in Patna;
Bhulia, in Patna.

NORTHERN ZONE (PAHARI, HIMALAYAN)

The four languages of the Northern Zone are spoken by at least eight million people in Nepal and in the most northwestern part of India on the slopes of the Himalayas in the states of Punjab and Himanchal Pradesh. The most widely known language in the Northern Zone is Nepali; but only some of the eight million Nepali speakers speak Nepali as a native language.

The other three languages of this zone (Kumoan, Garhwali and Western Pahari) are each spoken by at least one million people.

(19) Nepali (Nepalese, Gorkhali, Khas Kura, Parbatiya, Eastern Pahari) is spoken as a lingua franca and used as a vehicle of literature and administration by most of Nepal's inhabitants. It is the native language of the Gurkhas and tribes of Nepal west of the Kali River; it is spoken as a native language in urban centers, and in the lower Himalayas. It is less used as a native language east of the Kali River, where Nepali competes with Sino-Tibetan languages. In addition to the adjacent Sino-Tibetan languages, Nepali is bounded on the southeast by Bihari and by Eastern and Western Hindi on the southwest. To the northwest towards Tammu and Kashmir, Nepali is bounded by Kumaoni and Garhwali; at the Indo-Nepalese western frontier, Nepali dialects 'merge' into Kumaoni and Garhwali. No information is available on Nepali dialects; some dialect differentiation would be expectable, since there are many enclaves of Nepali speakers in the isolated reaches of Nepal.

D. R. Turner, Dictionary of Nepali Language, 1931, in discussing the close resemblances of Nepali to Rajasthani, concludes that 'Nepali ... appears to have belonged originally to a dialect group which included the ancestors of Gujarati, Sindhi, Panjabi, and Hindi [West Hindi]'; the special features it has in common with Rajasthani is due to the preservation of common original features rather than the introduction of common innovations. The existence of certain Bihari-like features in Nepali permits

Turner to postulate a pre-Nepali Indic language in Nepal with Bihari affinities.

Western Hindi is wielding an increasing influence on the Nepali vernacular through the Hindusthani dialect of Western Hindi; and on the Nepali literary dialect (through the prestige of High Hindi).

(20) Kumauni (Kumaoni) is spoken by a million people, primarily in the district of Almora and the northern portions of Naintal District in northwestern India (from the border of Nepal westward). It borders Nepali to the east, Garhwali to the west, Western Hindi to the south, and Sino-Tibetan languages to the north. With Garhwali, Kumauni is classified in the Central Pahari subgroup of the Northern group of Indic languages.

Thirteen Kumaoni dialects are listed by Grierson:

Khasparjiya, spoken in the center of Kumaun, in Pargana Barahmandal and in the adjoining parts of Pargana Danpur, both in the Almora District;

Phaldakotiya, spoken in Pargana Phaldakot of Almora and the north of Naini Tal District;

Pachhai, spoken in the southwest of Almora, on the borders of British Garhwal, and immediately to the west of the above two dialects;

Kumauni, in Naini Tal;

Bhabari, in Rampur;

Kumaiya, in Kali Kumaun just east of Naini Tal; Chaugarkhiya, in the Pargana of the same name;

Gangola, in the Pargana of that name;

Danpuriya, in the Pargana of Danpur;

Soriyali, in the Pargana of Soriyali;
Askoti, in the Pargana of Askoti;
Sirali, in the Pargana Sirali;
Johari, in the Pargana of Johar.

(21) Garhwali is spoken by about 1,200,000 people in the Tehri-Garhwal and Garhwal Districts of India west of Kumauni, southeast of Western Pahari, east of Panjabi, north of Western Hindi. With Kumauni, Garhwali is classified in the Central Pahari subgroup of the Northern group of Indic languages.

There are nine dialects of Garhwali listed by Grierson; all (except Tehri-Garhwal) are spoken in the Garhwali District in addition to the areas specified below:

Srinagariya (Standard);
Rathi (Rathwali), also a few speakers in Almora;
Lohbya, also a few speakers in Almora;
Badhani;
Dasaulya;
Majh-Kumaiya, also in Almora;
Nagpuriya;
Salani, also spoken in Almora, Dehra Dun, and by a few people in Saharanpur, Bijnor, and Moradabad;
Tehri Garhwal, spoken in Tehri State.

(22) Western Pahari is spoken by nearly two million people in the

States of Himachal Pradesh, Punjab and Uttar Pradesh. It is spoken in the districts of Sirmoor, Chamba, Mandi, and Mahasu in Himachal Pradesh, Kangra and Simla in the Himalayan Division of Punjab; and in the northern corner of Dehra Dun in Uttar Pradesh. Western Pahari is bordered on the southwest by Garhwal, on the south by West Hindi and Panjabi, on the west and northwest by Dardic languages and on the north and northeast by Sino-Tibetan languages. There are a great number of dialects in Western Pahari—twenty-two have been reported; Grierson lists the following nine dialects or groups of dialects:

Jaunsari
Sirmauri
Baghati
Kiuthali
Satlaj Group (a small set of dialects)
Kului
Mandeali
Chameali (4 subdialects)
Bradrawah (3 subdialects).

UNCLASSIFIED GYPSY DIALECTS

A number of wandering 'gypsy' (but not Romany, the extra-Indian Gypsies) tribes and castes are distinguished in India. Their languages have been little studied, but from what information is available they seem to be mostly dialects of the languages spoken by the more stable populations of the areas they inhabit; some, however, as Dumaki, may represent separate languages. Grierson reported half a million speakers of such dialects, but the 1951 census shows only about 2,500. Grierson's list of dialects and 'argots' spoken by gypsy groups includes the following among Indic dialects which are not assigned to specific Indic languages.

Dumaki (Doma) is spoken by just over 300 people in Hunza and Nagar among the Burushaski. In Hunza they are concentrated (by the dominant Burushaski) in an allotted area known as Berishal, one mile from Batut, the capital. Dumaki affinities with Romany have been suggested on the basis of vocabulary similarities. Dumaki does not possess close relations with the neighboring Dardic languages nor with Burushaski, which serve as donors for the large borrowed vocabulary of Dumaki.

Pendhari is spoken only as a 'home language' by the bilingual Pendhari who speak the Dakhini Hindustani dialect of Western Hindi as a second language. Linguistically Pendhari shares features with Dakhini Hindustani, with the Jaipuri dialect of Rajasthani, and with Marathi. The census of 1911 reported 6,413 Pendhari speakers; it is noted that many bilingual Pendhari were probably counted as Hindustani speakers. According to the information of the 1911 census, the Pendhari live in Belgaum and Dharwar Districts of Mysore and also in Madhya Pradesh; no separate report of Pendhari speakers was given in the 1951 census.

Kolhati (Bhatoo, Doomun, Kollati, Dombari, Dombhari) is spoken by a tribe of rope dancers and tumblers living primarily in Maharashtra and Madhya Pradesh. The Kolhati numbered over 12,000 in the 1911 census; however, in the 1951 census, only 995 were counted. Linguistically, Kolhati shares features with Sāsi and Panjabi (and also to some extent with Gujarati). The Kolhati also have an argot based on systematic alteration to disguise their language.

Sāsi (Sāsiya, Bhattu) is spoken by a migratory 'criminal tribe' in Panjab and Uttar Pradesh who speak a language which isclosely related to Hindustani and Panjabi, and said to be intermediary between the two. The Sāsi have an argot, which they call Farsi, based on changing individual words of their ordinary dialect so as to make them unrecognizable.

Beldari is spoken by some members of a group called the Beldar who live in the Maharashtra Districts of Amravati, Buldana, Thana, North Satara, South Satara and Kolhapur, the Mysore District of Belgaum, and in the Rajasthan District of Jaisālmer. The 1911 census reports that over 5,000 persons reported their native language to be Beldari, but that additional speakers of the language seem probable, since all Beldari speakers are bilingual in the predominant language of the region where they live. The 1951 census reports 3,853 speakers of Beldari in Central and Western India. Linguistically, Beldari shares many features with Eastern Rajasthani and also some features with Marathi and Oriya.

Garodi (Garudi) is spoken by the Garodi, a wandering tribe of jugglers in the Belgaum District of Mysore. In 1951 only 17 speakers were reported for Garodi, although others may have given their second language for the census reports. Garodi is said to show close affinity with Hindustani, Marathi, and Rajasthani.

Myanwale (Lhari, Lohari) is spoken by at least some, if not all, of the Lohar in the Belgaum District of Mysore. In 1911 there were 817 Lohars reported; only 105 were reported for the 1951 census. Their language shows closest affinity to Dakhani Hindustani, and Rajasthani and Gujarati.

Sikalgari (Saiqatri, Siqligari) is spoken by the Sikigar, a caste of armorers and polishers of metal. In 1911 there were around 6,000 reported in Rajasthan by the census; the 1951 census reports only 584 as speakers of Sikalgari. The closest affinity of Sikalgari is with Gujarati but it also shares features with Panjabi, Rajasthani and Bhili.

Kanjari (Kuchbandi) is spoken by the Kanjar, an aggregate of vagrant tribes, located primarily in Uttar Pradesh, Rajasthan, Maharashtra, and Mysore. Grierson reports that there were over 7,000 speakers of the language in 1911; however, in the 1951 census there were only approximately 5,000 speakers reported. Kanjari shares features with Western Pahari, Rajasthan, Panjabi, and Gujarati. The Kanjars also have an argot based on their regular language that is formed by systematic alterations. Grierson reports that it is impossible to separate Kanjari from the closely similar Dom. Dom (Domra Magu hiya) is spoken by some members of a menial caste who live primarily in Assam, West Bengal, Bihar, Uttar Pradesh, Panjab Madhya Pradesh, Jamanu Kashmir, and Orissa. Grierson describes Dom as an argot of the Bhojpuri dialect of Bihari that shares some features with Rajasthani. Although Grierson gives 128,500 as the total number of Doms, he notes that not nearly all of this number speak Dom. The 1951 census lists 24 people as speaking Dom.

POSTSCRIPT FOR INDIC

On the vexed question of linguistic states, it may be said for Indic languages (as indeed it is said, above) that none are in the offing; or it may be said that all reallocations of the political boundaries have been in response to an expressed public opinion in favor of linguistic states, even though the redrawn boundaries never seem to have the desired result of including in one political state all the dialects of one language while excluding dialects of other languages; or it may be said, even more strangely, that linguistic states in India (states with predominant languages) are already in existence and may be counted as numerous, including Orissa, Mysore, and many others.

These different ways of predicting, or stating the modern development, or the present realization of linguistic states are only in seeming conflict. They reflect, at bottom, attempts to formulate some ordinary expression to account for an extraordinary linguistic situation--one occasioned by the existence of an enormous number of dialects, both of the vertical and horizontal types. They reflect the fact, acknowledged by all Indic specialists, that it is rare to be able to point to a clear-cut boundary between one language and another (where political boundary lines might be drawn); instances of such exceptional segregation of one language from another have been carefully noted above. In general, however, the opposite holds true--a situation often described in terms of transitional dialects between languages. And between such languages — between the majority of language centers — buffer areas would have to be recognized, if political states were to closely approximate linguistic states.

The general situation, as Murray Emeneau has said, reflects less sharp distinctions among Indic languages than exist between Polish and Russian among Slavic languages. And the distinction between Polish and Russian, as Edward Stankiewicz has said (in conversation), seems relatively clear-cut only because a political border separates the two nations; it does not seem clear-cut to Slavic dialectologists who look for the boundary between Polish and Russian along the border.

The modern interlanguage ecology in India and Pakistan is not restricted to closely related languages in one branch or sub-branch, as Indic, but is so extensive as to involve languages belonging to different language families, as Dravidian and Indo-European. Under Sinhalese-Maldivian, above, it was noted that there might be examples of Dravidianization of some Indic languages; for the latest published instance of such Dravidianization in syntax, see BSOAS 27. 129-50, especially 137 ff (1964).

And this interlanguage ecology, as well as the intralanguage ecology, occasioned by the enormous dialect differentiation of given languages, is not merely a modern development but one which has great historical depth, as we are reminded in a still unpublished paper by Murray B. Emeneau ('... a paper that is to appear as part of the UCLA Indo-European conference of a year ago'). The paper gives the dialect evidence and interpretation obtainable from classical Sanskrit (circa 500 B.C.) and the earlier period of Vedic Sanskrit, when the hymns of the Ṛgveda were composed (circa 1200-1000 B.C.).

"Classical Sanskrit is a literary language written according to the book, i.e. Pāṇini's grammar, and following it more or less correctly. We find in it no dialects, no chronological development, except loss and at times invasion from the vernaculars of the users, and no geographical divergences. Vedic Sanskrit, however, is ... different. It is anything but a unified language, a language of one dialect only. It shows even within the oldest member of the corpus, the Ṛgveda, linguistic features that can be explained only by positing their origin in slightly differing dialects, and within the total Vedic corpus there is a sliding scale of clusters of dialectal features that run all the way from those that are most different from classical Sanskrit to those that are, in fact, taken by most scholars in the field to be essentially the dialect that Pāṇini described as his norm." (ms. p. 1).

After marshalling linguistic evidence, the relationship between Vedic and classical Sanskrit is stated.

"The Ṛgvedic dialect, then, is clearly not the direct ancestor of classical Sanskrit. There must have been, even on this much evidence, several closely related dialects in the period of the Ṛgveda composition, one of which is the basic dialect of this text, another of which is basically the ancestor of the classical language of some centuries later. But it is also clear that the Ṛgvedic linguistic norm, even apart from hymns that represent something very close to the classical language, was a mixed dialect, and that one of the elements in the mixture was something near to classical Sanskrit." (ms. pp. 7, 8).

Post-classical evidence is next reviewed. However, these later dialects and literary languages do not necessarily give additional information on the speech of Indo-European speakers, who migrated to India; or, on the other hand, they may shed light on the speech of these early immigrants, and hence add to the reconstructions of Indo-European.

"If the former linguistic hypothesis were the case, the picture would be like that of the Romance languages coming from IE through the Latin channel only and adding nothing to our knowledge of IE that we do not already get from the Latin record--this at least seems to be implicitly, or even more or less explicitly, the standard doctrine about the Romance languages, though it is at least possible that the picture is overdrawn and that there are a few scraps of evidence for IE to be extracted from the Romance languages (so Malkiel in conversation). The other case would be more like that of Germanic, where the literatures of the medieval period do not remotely exhaust all the languages and dialects of that period and where even scraps of other medieval evidence and much of the modern material is employable for IE research." (ms. pp. 14, 15).

The possible contributions to this question offered by the Dardic languages (12, below) is next discussed.

"It was claimed by Grierson (as well as by some before him) that these two groups of languages [Dardic and Kafir] form a third branch of Indo-Iranian, in that 'they seem to have left the parent stem after the Indo-Aryan languages, but before all the typical Iranian characteristics, which we meet in the Avesta, had become developed.' The material which was gathered by Morgenstierne after Grierson's volume appeared, led Morgenstierne to the conclusion (which has been accepted by, e.g., Jules Bloch and Burrow) that the Dardic languages

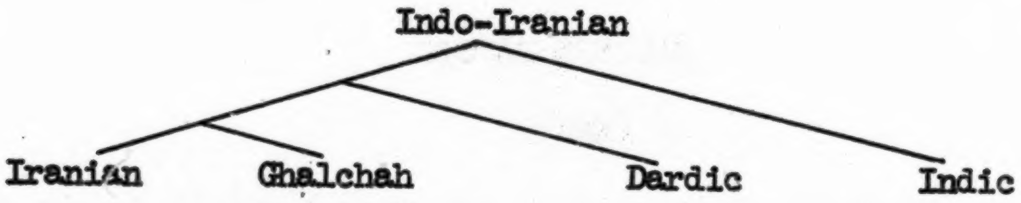
(Kashmiri, Shina, Indus Kohistani, Khowar, Kalasha, Pashai, Tirahi) are Indo-Aryan but did not pass through the MIA developments represented by the records, while, on the other hand, the Kafir languages (Kati, Waigali, Ashkun, Prasun, and to some extent Dameli) may occupy some sort of special position. The task of sorting out the evidence is considerably complicated by loanwords in the Kafir languages from neighboring Iranian languages and from other neighboring Indo-Aryan languages, and also by loans in the other directions, i.e. from the Kafir languages into neighboring Iranian and Indo-Aryan languages." (ms. pp. 20, 21).

THE DARDIC BRANCH OR SUB-BRANCH OF INDO-EUROPEAN

A score of Dardic (Pisacha) languages are spoken in Afghanistan and adjacent West Pakistan and India. The Dardic linguistic area extends south of the Hindu Kush Mountains from Nuristan (formerly Kafiristan) in Afghanistan in the west, to the Chitral country and the Indus and Swat Kohistans of extreme northern West Pakistan in the center, and to the Jammu and Kashmir area on both sides of the U.N. Cease Fire Line between India and Pakistan in the east. The conventional name for the region inhabited by Dardic speakers (excluding Afghanistan) is Dardistan. Kashmiri is the only Dardic language with a literary tradition. The Dardic languages possess, in almost unaltered form, words which in India are seldom found except in Vedic Sanskrit. These words are in common use in Dardic. In each Dardic language there is also a small element of Burushaski spoken immediately to the northeast of Shina (but Burushaski is neither Indo-European nor Dravidian).

The linguistic neighbors of the Dardic languages are Pashto (Iranian), the Ghalchah languages (also Iranian), the uniquely unrelated Burushaski language, several Sino-Tibetan languages, and a few Indic languages: Pahari, Lahnda and Punjabi. The geographic position of these languages vis-à-vis the Dardic languages is west and southwest (Pashto), northwest (Ghalchah), northeast (Burushaski and Sino-Tibetan), southeast (Pahari), and south (Lahnda and Punjabi).

The Indo-European branch affiliation of the Dardic languages is clear, but not entirely so. There is apparent concensus in the conclusion of two recent scholars (Morgenstierne, Emeneau): Dardic belongs in the Indic branch of Indo-European. The view of earlier scholarship (Grierson and others) was that the Dardic languages belong in the Iranian branch of Indo-European—an offshoot of Iranian—as shown in the following chart:



A third view on the classification of Dardic places these languages in a third branch of the Indo-European family, beside Iranian and Indic, in South Asia. This view is maintained, for example, by Muhammad Shuja Namus, Origin of Shina Language, Pakistani Linguistics (1962), Anwar S. Dil, ed., 1963, pp. 55-60.

Population figures for most of the Dardic languages are not available. Kashmiri may have several millions of speakers; Shina 100,000; and the rest not more than tens of thousands for each language. A reasonable total would be less than a half million (if Kashmiri were excluded).

The list of Dardic languages below, is divided into the three major groups given by all investigators:

Western Dardic (Kafir)
Central Dardic (Khowar)
Eastern Dardic (Dard).

The group or branch names appear as center heads below followed by a discussion of the language-dialect problem for each. The boundary between language and dialect is not always determined, but it is still possible to indicate clearly the relative magnitude of diversity for each of the three groups. There are a dozen languages in the Western Dardic (Kafir) group, but a single language constitutes the Central Dardic (Khowar) group. There are a half dozen languages in the Eastern Dardic (Dard) group—more or less, depending on whether the languages counted after Phalura are separate languages or dialects of Phalura.

WESTERN DARDIC

(KAFIR)

The dozen languages of this group are spoken mainly in the Nuristan (Kafiristan) region of Afghanistan north of the Kabul River and in adjoining West Pakistan. One of the Kafir languages, Tirahi, is separated from the others, being located south of the Kabul River west of the Khyber Pass. Tirahi is entirely surrounded by Pashto (Iranian) speakers.

(1) Bashgali (Kati) is spoken on the Afghan-Pakistan border north of the Bashgal River and west of the Kunar. A fairly divergent dialect of Bashgali is spoken by settlers from Kamdesh in Lower Chitral and known locally as Shekhani.

Other dialects of Bashgali, beside Shekhani, are:
Western Kati
Badimuk
Bargromatal.

(2) Wai-ala (Wai).

(3) Wasi-veri (Veron); languages (2) and (3) are spoken in Afghanistan next to Bashgali (1).

(4) Ashkund is spoken north of Pashai in Afghanistan.

(5) Kalasha is spoken in a few villages along the Kunar River, southwest of Drosh, Pakistan.

(6) Gawar-bati (Narsati) is spoken in Southern Chitral, south of Kalasha, on the left bank of the Kunar River and in several villages on the Afghan side of the border. The Shumasht dialect is heavily influenced by Pashai; this

dialect is isolated from the others. Shumasht speakers live 60 miles further up the Kunar River on the Chitral frontier.

(7) Pashai (Laghmani, Deghani) is spoken in Afghanistan between Nuristan and the Kabul River. More specifically, Pashai is spoken from the Kunar Valley in the east across various tributaries of the Kabul River to the Panjshir Valley in the northwest. According to Georg Morgenstierne (Indo-Iranian Frontier Languages, Vol. 3, Oslo, 1944) Pashai is split into a large number of mutually incomprehensible dialects, namely:

Gulbahar
Chilas
Aret
Wegal
Darrai Nur
Laurowan.

Morgenstierne leaves no doubt that '...in spite of all dialectal differences... Pashai is decidedly one language, well defined through phonetical, and especially through morphological and lexical peculiarities' (Report on a Linguistic Mission to North-Western India, Det Mallingske Bogtrykkeri, Oslo, 1932, p.24).

(8) Bashkarik is spoken in several villages in Dir Kohistan. Diri, mentioned by Grierson, is a form of Bashkarik.

(9) Tirahi is spoken in a few Afghanistan villages southeast of Jalalabad and west of the Khyber Pass. Tirahi, of all the Kafir languages, is most closely related to Kohistani.

(10) Prasun may possibly be a separate language; if not, it is a divergent dialect of Bashgali (1). Prasun is spoken in Chitral in the

villages of Usut, Zumu and Saici. It is very closely related to Bashgali, but is more archaic. It is probable that Prasun is a remnant of a very ancient population (Morgenstierne, 1932). Prasun has more influence from Iranian than any other language in the Western Dardic (Kafir) group.

Other languages or dialects in this group are:

(11) Gujuri, spoken in a few hamlets in the Shishi Valley and elsewhere in Chitral;

(12) Waigeli and Zhonjigali(with corresponding village names).

CENTRAL DARDIC

(KHOWAR)

The Khowar language (Chitrali, Chatrari, Arniya) is the sole member of this group. Of the eleven languages of Chitral, Khowar is the most important. Due to recent dispersal of Khowar speakers, there are as yet no very pronounced dialectal variations in the Khowar language. Grierson believes Khowar to be very closely related to an Iranian group of languages known as the Ghalchah languages which are spoken immediately northwest of the area occupied by Khowar speakers. Morgenstierne views Khowar as clearly Indic, despite its being flanked by Iranian languages. There is a non-Indic as well as non-Iranian component in Khowar; it contains a considerable number of lexical items of unknown etymology which are apparently neither Indic nor Iranian, but possibly Burushaski which is also neither Indic nor Iranian. But Khowar is at least an Indo-European language, while Burushaski is not. Khowar is number (13) in our list of Dardic languages.

EASTERN DARDIC

(DARD)

The half dozen languages [illegible] in this final group of Dardic languages are found in a wide area in eastern Peshawar in Pakistan and in Jammu and Kashmir on both sides of the U.N. Cease Fire Line. Kashmiri, the easternmost language of this group, has been estimated to have as few as one and a half million speakers and as many as three or four million speakers.

(14) Shina, with perhaps as many as 100,000 speakers, is spoken in Jammu and Kashmir from the Afghan and Pakistan borders to the U.N. Cease Fire Line. The main dialects listed below are those given by T. Grahame Bailey (Grammar of the Shina Language, Royal Asiatic Society, London, 1924):

Gilgiti
Astori
Kohistani.

From Grierson the Bailey list can be expanded to include:

Chilasi
Gurezi
Brokpa.

On the north, the Shina dialects are bordered by Burushaski; on the northeast and east, by Balti and Ladakhi (Sino-Tibetan languages); and elsewhere, by other Dardic languages.

(15) Kashmiri is spoken in Jammu and Kashmir on both sides (but mostly south) of the U.N. Cease Fire Line, south of the area occupied by the Shina (14). In the southwest, Kashmiri is bordered by Lahnda; in the south, by Panjabi speakers; and in the southeast, by Pahari speakers. Virtually the only information we have concerning Kashmiri dialects comes from Grierson who lists three Kashmiri dialects:

Standard Kashmiri

Kashtawari

Transitional (to Punjabi).

(16) Kohistani is spoken mainly in the Swat Kohistan of West Pakistan. The number of speakers, reported to be 7,000 by Suniti Kumar Chatterji (Indo-Aryan and Hindi, Firma K.L. Mukhopadhyay, Calcutta, 1960), seems to be too low. There are three Kohistani dialects (possibly separate languages):

Torwali

Garwi

Maiya (of Mayo District, Indus Kohistan).

(17) Phalura (Palula) is spoken by 800 to 1,000 speakers in a few villages in side valleys on the eastern side of the lower Chitral Valley. This language has been discovered since Grierson's Linguistic Survey of India. One of the interesting phonological features of the Dardic languages generally is deaspiration of voiced stops — a development that is presumably recent since it is still incomplete in some Dardic languages. Phalura is in this respect more archaic than the other Dardic languages, since it maintains a series of aspirated voiced stops, recently interpreted as clusters of stop plus /h/ by Georg Morgenstierne (Notes on Phalura, Videnskaps-Akademi i, Oslo, 1941).

Three languages closely related to Phalura are spoken in:

(18) Sau (on the Kunar River in Afghanistan);

(19) Ashret (in Dir Kohistan, Pakistan);

(20) Damel (in the Gid Valley on the east side of the Kunar River in Southern Chitral, Pakistan).

Phalura (17), together with the unnamed languages numbered (18), (19) and (20), has been called Dangarik. Further information is expected to show that (18), (19), and (20) are not separate languages, coordinate with Phalura, but rather, with Phalura, divergent dialects of a single Dangarik language.

DARDIC SOUND SYSTEMS

Pashai (Laurowani dialect, after Morgenstierne, 1944) and Kohistani (Torwali dialect, after George A. Grierson, Torwali, an Account of a Dardic Language of the Swat Kohistan, Royal Asiatic Society Prize Publication Vol. 9, London, 1929) have relatively simple consonant (especially stop) systems:

p	t	ṭ	č	k	
pʻ	tʻ	ṭʻ	čʻ	kʻ	
b	d	ḍ	ǯ	g	
	s		š	x	h
	z		ž	γ	
m	n				
	l				
	r	ṛ			
w			y		

The above system is exactly that of Torwali (excluding sporadically occurring voiced aspirated stops which are being replaced by unaspirated ones). For Pashai add /ṣ ẓ/.

The most elaborate consonant system of all the Dardic languages is to be found in Phalura (Morgenstierne, 1941):

p	t	c	ṭ	c̣	č	k
pʻ	tʻ	cʻ	ṭʻ	c̣ʻ	čʻ	kʻ
b	d	ʒ	ḍ	ʒ̣	ǯ	g

	s	ṣ	š	x	h
	z		ž	γ	
m	n	ṇ			
	l				
	r	ṛ			
w			y		

In addition one should mention the aspirated voiced stops (and affricate) considered to be clusters of stop + / h/ by Morgenstierne due to an intervening phonetic vowel quality. These clusters include / bh dh ḍh ǯh gh/ .

The consonant system of Gawar-bati is almost as diversified as that of Phalura, lacking only / ʒ̣ ẓ/ but adding / L/ , described as a 'pure unvoiced l' by Georg Morgenstierne (Notes on Gawar-bati, Videnskaps-Akademi i, Oslo, 1950). Also, in Gawar-bati, voiced aspirated stops are rapidly being replaced by unaspirated ones.

The Shina consonant system (after Bailey, 1924) differs from Phalura and Gawar-bati only slightly:

p	t	ṭ		c̣	č	k	
p'	t'	ṭ'		c̣'	č'	k'	
b	d	ḍ		ʒ̣	ǯ	g	
f	s		ṣ		š		h
v	z		ẓ		ž		
m	n		ṇ				
	l						
	r		ṛ				
					y		

The Kashmiri consonant system (the dialect of the educated people of the city of Srinagar, as described by Ashok R. Kelkar and Pran Nath Trisal, Kashmiri Word Phonology—*AL* 6.1.13-22, 1964) differs from other Dardic languages chiefly in that it contains only three fricatives:

p	t	c	ṭ	č	k	
p·	t·	c·	ṭ·	č·	k·	
b	d	z̧	ḍ	ž̧	g	
		s		š		h
m		n				
		l				
		r				
w		H		y		

/H/ is described as a high central unrounded semivowel.

Like the consonant systems, the vowel systems of Dardic languages are fairly homogeneous, ranging from the 2 (FB) over N type, /i e a o u/, plus length (and in some cases also infrequent nasalization) in Gawar-bati, Kohistani and Phalura, to a 2 (FCB) over N type, /i e ɨ ə a o u/, plus length in Kashmiri.

A different seven vowel system occurs in Bashkarik (Morgenstierne, 1941):

i		u
e	ə̈	o
æ		a

The exact vowel system of Shina has not been clearly stated, but there is some evidence for phonemic tone in this language, as witness lēl *blood* (with level tone), contrasting with lĕ́l *visible*, (with low rising tone).

The Following Abbreviations Will Be Used

AA	American Anthropologist
ACLS	American Council of Learned Societies
AES-P	American Ethnological Society, Publication
AL	Anthropological Linguistics
APS-P	American Philosophical Society, Proceedings
APS-T	American Philosophical Society, Transactions
BAE-B	Bureau of American Ethnology, Bulletin
BAE-R	Bureau of American Ethnology, Report
CU	Columbia University Contributions to Anthropology
IJAL	International Journal of American Linguistics
IUPAL	Indiana University Publications in Anthropology and Linguistics
JAF	Journal of American Folklore
JSAP	Journal de la Société des Américanistes de Paris
Lg	Language
RCPAFL	Research Center Publications in Anthropology, Folklore and Linguistics
SJA	Southwestern Journal of Anthropology
SIL	Studies in Linguistics
TCLP	Travaux du Cercle Linguistique de Prague
UMPL	University of Michigan Publications, Linguistics
UCPAAE	University of California Publications in American Archaeology and Ethnology
UCPL	University of California Publications in Linguistics
VFPA	Viking Fund Publications in Anthropology
WDWLS	William Dwight Whitney Linguistic Series

CONTENTS

INDO-EUROPEAN FASCICLE ONE

1. Branches of the Indo-European family 2
2. Italic 11
3. Germanic 43
4. Celtic 90
5. Slavic 100
6. Baltic 152
7. Greek 158
8. Albanian 177
9. Armenian 181
10. Iranian 190
11. Indic 219
12. Dardic 284